

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1709 <TIGR>
A:Cross-references: GB:032710; GB:L42023; NID:g1573200; PID:g1573211; TIGR:HT0248

Alignment_scores:

Quality:	2541.00	Length:	701
Ratio:	4.092	Gaps:	5
Percent Similarity:	88.368	Percent Identity:	67.618

Alignment_block:

US-09-303-518d-649 x C64057

Align seg 1/1 to: C64057 from: 1 to: 709

```
79 TACTTAGCCATATGCTGCTGCGCATTCCTCCCAAGCCGTGGGGG 128
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
10 PheLeuThrAlaCysValSerLeuGlyLeuAlaSerGlnAlaTrpAlaG1 26
129 ACACACTATTTCGGCATCAACTACCAATACATCCGACCTTCCGAAA 178
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
26 YHisThrTyrPheGlyIleAspTyrGlnTyrTyrArgAspPheAlaG1ua 43
179 ATAAAGCAGATTGTCAGTCGGGGGAAAGATATTGAGTTTACACAAA 228
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
43 snLysGlyLysPheThrValAlaAlaLysasnIleGluValTyrAsnLys 59
229 AAAGGGAGTTGGTCGGCAAAATCAATGACAAAAGCCCGATGATTGATT 278
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
60 GluGlyGlnLeuValGlyThrSerMetThrLysAlaProMetIleAspPh 76
279 TTCTGTGCTGTCGGCTACGCGCGTGGCGGCGATTTGGCGCATATATA 328
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
76 eSerValValSerArgasnGlyValAlaAlaLeuValGlyLyspGlnTyrI 93
329 TTGTGAGCGTGGCACATTAACGGCGCTTAAACAACGTGATTGTTGGTGG 378
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
93 LeValSerValAlaHisasnGlyLysTyrTrpAsnAspValAspPheGlyAla 109
379 GAAGGAGAAATCCCGATCAACATCGCTTTACTTATTAATTTGGAACG 428
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
110 GluGlyArgasnProAspGlnHisArgPheThrTyrGlnIleValLysAr 126
429 GAATATATATAAGCAGGAGACTAAAGGCACTTATGCGGCGCATATATC 478
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
126 gAsnsmTyrGlnAlaTrpGluArgLysHisProTyrAspGlyAspTyrH 143
479 ATATCCCGCTTTGCATTAATTTGTACAGATGACAGAACTGTGGAATG 528
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
143 IsMetProArgLeuHisLysPheValThrGluAlaGluProValGlyMet 159
529 ACCACTTATATGATGGCGGCAATATATCATCAAAATTAATACCTCGA 578
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
160 ThrThrAsnMetAspGlyLysValTyrAlaAspArgGluAsnTyrProG1 176
579 CCGTGTGCTATTTGGCGAGGAGCATATTTGGCATCTGATCGAAGATG 628
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
176 uArgValArgGlyIleGlySerGlyArgGlnTyrTrpArgThrAspLysAspG 193
629 AGCCCAATTAACCGGCAAGATTCATATCATATTCGAAGTCCGCTATTCCTG 678
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
193 lGluIuThrAsnValHisSerSerTyrTyrValSerGlyAlaTyrArgTyr 209
679 CTCGTGTGCTGCAATACCTTTGCACAAATATGATCAGTGTGGCAGACT 728
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
210 LeuThrAlaGlyAsnThrHisThrGlnSerGlyAsnGlyLysThrIva 226
729 CAACTTAGTGTGTAAGAAATTAACATATGACCATATGTTTATACCAA 778
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
226 lAsnLeuSerGlyAsnValValSerProAsnHisTyrGlyProLeuProT 243
779 CAGGAGGCTCATTTGGCGAGAGTGGCTCACCAATGTTTATCTATGATGCC 828
```

```
||||| 259 hrcLysLysSerLysLysAspSerGlySerProMetPheThrIleTyrAspAla
829 CAANAAGCAAAAGTGTATTTAATGGCGTATTCGAACAGCGCAACCCCTA 878
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
260 LysLysLysGlnTyrPheLeuIleAsnAlaValLeuGlnThrGlyHisProPh 276
879 TATAGAAAGAAAGCAATGGCTTCAGCTGGTTCGTAAAGATTGTTCTATG 928
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
276 ePheGlyArgGlyAsnGlyPheGlnLeuIleArgGlnGluTyrPheTyrA 293
929 ATGAAATCTTGTCTGAGATACCATTCAGTATTC.....TAGAACCA 972
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
293 snGluValLeuAlaValAspThrProSerValPheGlnArgTyrIlePro 309
973 CGTCAAAATGCGAAATACCTTTTAAACGACGATTAATAGCAGCAAGAAA 1022
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
310 ProIleAsnGlyHisTyrSerPheValSerAsnAsnAspGlyThrGlyLys 326
1023 AATCAAT...GCCAAACATGAACACAAATTCCTGCTTAAATGATTAAAAA 1069
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
326 sLeuThrLeuThrArgProSerLysAspGlySerLysAlaLysSerGluV 343
1070 CACGAACCGTCAATTTGTTTAAATGTTTCTTAAATCCGACACAGACAGA 1119
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
343 alGlyThrValLysLeuPheAsnProSerLeuAsnGlnThrAlaLysGlu 359
1120 CCTGTTTATCATGCTCAGCTGAGTGTCAACAGATTATCCAGCCACAGCTGA 1169
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
360 HisVal...LysAlaAlaAlaGlyTyrAsnIleTyrGlnProArgMetG1 375
1170 TAATGAGAAATATATTCCTTATTAATGACGAGAAAGCGAATTGATAC 1219
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
375 uTyrGlyLysAsnIleTyrLeuGlyAspGlnGlyLysGlyThrLeuThrI 392
1220 TTACACGACAAATCAATCAAGGTCTGAGAGATTATATTCACAGACAT 1269
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
392 lGluLysAsnIleAsnGlnGlyAlaGlyLysLeuTyrPheGlnGlyAsn 408
1270 TTATACGCTC...TCGCGTGAATAATACGAAACTTGGCAGGCGGCGCT 1316
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
409 PheValValLysGlyLysGlnAsnAsnIleThrTrpGlnGlyAlaGlyVa 425
1317 TCATATCAGTAGAAGACAGTACCGTTTACTTGAAGATTAACGGCGTGGCA 1366
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
425 lSerIleGlyGlnAspAlaThrValGluTrpLysValHisAsnProGluA 442
1367 AGCAGCGCTGTCCAAATCGCAAGGCAACGCTGCACGTTCAAGCCAAA 1416
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
442 snAspArgLeuSerLysIleGlyIleGlyThrLeuLeuValAsnGlyLys 458
1417 GGGGAAACCAAGGCTCGATACAGCTGGGCGAGCTACAGTATTTGGA 1466
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
459 GlyLysAsnLeuGlySerLeuSerAlaGlyAsnGlyLysValIleLeuAs 475
1467 TCAGACAGCGAGCATTAAGGCAAAACCAAGCCTTTAGTGAATCGCT 1516
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
475 pGlnGlnAlaAspLualaglyGlnLysGlnAlaPheLysGluValGlyI 492
1517 TGCTACGCGGCGAGGGTACGGTCAACTGAATGCCATATACGTTCACAC 1566
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
492 lValSerGlyArgAlaThrValGlnLeuAsnSerThrAspGlnValAsp 508
1567 CCGGACAACTCTATTTGCGCTTCCGCGGAGCGTGTGATTTAAACGG 1616
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
509 ProAsnAsnIleTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnG1 525
1617 GCATTCGCTTGTTCACCGTATTCAAATACCATGAGAGGGCGCATGA 1666
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
525 YHisSerLeuThrPheLysArgGlyIleGlnAsnThrAspGlnGlyAlaMetI 542
1667 TTGTCAACCAATTCAAAGCAAGATTCACCGCTTCCATTACAGGCAT 1716
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

542 IeValAsnHisAsnThrThrglnValAlaAsnIleThrIleThrglyAsn 558
1717 AAGATATTGCTACACCGCAATTAAC.....AACACTGGGATAG 1757
559 GluSerIleThrAlaProSerAsnLysAsnIleAsnLysLeuAspTyr 575
1758 CAAAAAGAAATGGCTACACCGGTTGGTGGAGAAAGATACGACCA 1807
575 rSerLysGluIleAlaTyrAsnGlyTyrPheGlyGluThrAspLysAsnL 592
1808 AAACGACGAGGCGGCTCAACCTGTTTACGACCCCGCGGACGAAAGACCGC 1857
592 yShAsnGlyLysLeuAsnLeuIleTyrLysProThrThrgluAspArg 608
1858 ACCCTGCTGCTTCCGCGGCAACAAATTTAAACGCAACATCAGCAATC 1907
609 ThrLeuLeuSerGlyGlyThrAsnLeuLysGlyAspIleThrIle 625
1908 AAACGCAACTGTTTTCACGCGGACGACACACCGCAGCCGTCACATC 1957
625 rLysGlyLysLeuPheSerGlyArgProThrProHisAlaTyrAsn 642
1958 ATTAAAGACCATGTCGCGCAAAAAAGGCGCATTCCTCCGCGGGAATC 2007
642 ILeuAspLysArgTyrPheSerGluMetGluGlyIleProGlnGlyLeu 658
2008 GTGTGGGCAACGACTGGATCAACCGGACATTTAAAGCGAAACTCCA 2057
659 ValTrpAspTyrAspTyrPheAsnArgThrPheLysAlaGluAsnPheG 675
2058 AATTAAAGCGGACGCGGCTGTTTCCCGCAATGTTGCCAAAGTGAAG 2107
675 nIleLysGlyLysLeuAlaValAlaSerArgAsnValSerSerIleGlu 692
2108 GCGATGTCATTGAGCAATCAACCGGACGAGATTTTGTGTGCGACCG 2157
692 LysAlaThrThrValSerAsnAsnAlaAsnAlaThrPheGlyValValPro 708
2158 CAT 2160
709 Asn 709
seq_name: p1r2:S61314
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HPI3
C:date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:accession: S61314
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the Iga gene encoding Iga1 protease in Neisseria
A:reference number: S61314; MIMID:95302961
A:accession: S61314
A:status: preliminary
A:molecule type: DNA
A:residues: 1-1561 <LOW>
A:cross-references: EMBL:X82474; NID:9732873; PIDN:CAA57857.1; PID:9732874
C:superfamily: Iga-specific metalloendopeptidase
C:keywords: hydrolase, metalloproteinase
alignment_scores:
Quality: 1615.50 Length: 1753
Ratio: 1.766 Gaps: 55
Percent Similarity: 52.196 Percent Identity: 28.009
alignment_block:
us-09-303-518d-649 x S61314 ..
Align seg 1/1 to: S61314 from: 1 to: 1561

49 AAACGCGCGCATTCGCTTCGCTGCTTACTAGCATATGCTGTC 98
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeu 18
99 GTTCGGCATTCCTCCCAAGCCTGGGCGGACACACTTATTTTCGACATC 148
18 aTyrAlaLeuThrProTyrSerGluAlaLeuValArgAspAspVal 35
149 ACTAACCAATGCTATCGGACTTGTCCGAAATTAAGCAAGTTTGACATC 198
35 sPyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
199 GGGCGCAAGATATGTAGGTTTACACAAAAAAAGGAGAGTGGTGGCA 248
52 GylAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleLys 68
249 ATCAATGACAAAAGCCCGATGATGATTTTCTGTGTCGCGCATC 297
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsn 85
298 . GCGGTGGCGCATTCGCGGACATATATTTGTGACCGTGCACAT 345
85 ySalGylleAlaThrValAlaAspProGlnTyrAlaValSerValLys 101
346AACGGCGCTATACACGCTGA 368
102 AlaLysAlaGluValIleThrPheTyrTyrGlyGlnTyrAsnLys 118
369 TTTTGTGCGGGAAGAAAGAAATCCGATCAACATCGTTTACTATTA 418
118 nAspValAlaAspLysGluAsn.....GluTyrArgV 129
419 TGTGGAACGATATATTAATAA.....GCAGGAGCT 450
129 aValGluGlnAsnAsnTyrGluProHisLysAlaTrpSerIleAsn 145
451 AAAGGCAATCCTTATGCGCGCATATCAATGCGCGTTGCAATTAAT 500
146 LeuGlyArgLeu.....GluAspTyrAsnMetAlaArgPheAsnLys 160
501 TGTCAACAGATGCAAGACCTGTTGAATGACAGTTATATGATGGG..C 547
160 eValThrGluValAlaProIleAlaProThrAspAlaGlyGlyLeu 177
548 GGAATATATGATCAAAATTAATTCCTGACCGTGTGTTGGGCGCA 597
177 sPThrTyrLysAspLysAsnArgPheSerSerPheValArgValGlyAla 193
598 GCGAGCAAA.....TATTGGCGATCTGATGAAGATGACCCCAATTAACCG 641
194 GylArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluLys.. 209
642 CGAAAGTTCAAT.....CATATTCGAAGTGGATTTGCTGGCTGC 682
210 .GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAla 226
683 TTGTGCGC.....AATACG 696
226 IeAlaLeuThrProTyrLysAspIleAsnIleAspIleThrMetLeu 242
697 TTTGCACAAATGATCAGTGGTGGGACAGTCAACTTATGATGTAATA 746
243 GluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGlu 259
747 AATTAAACATAGC...CAATATGCTTTTAAACACAGAGCGTCATTTG 793
259 uLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeu 276
794 GCGACAGTGGCTCACCAATGTTATCTATGATGCCCAAAAGCAAGTGG 843
276 LysAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 292
844 TTAATTAAATGGGTAATTGCAACGCGCAACCCCTATATGAGAAAAAGCA 893

```

293 ValPheLeuGlyThrTyrAspTyrPalaGlyTyr...GlyLys.... 306
      ::::: ||::: :::: |||||
894 TGGCTTCAGCTGTTGCTAAAGATGGTCTATGATGAATCTTGCTG 943
      ||::: |||
307 .....LysSerTrpGlnGluTrpAsnIleTyr.... 315
      ||::: |||
944 GAGATACCATTCACTATCTACGACACGTCAAAATGGGAATACCT 993
      |||
316 .....LysLysGlu 318
      |||
994 TTTAACGAC.....GATATATGCGACAGGAAATCA 1028
      ||| ||| ||||| |||||
319 PheAlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGly 335
      ||| ||| |||
1029 TGCCAAACATGACACAACTCTGCTCAATAGATTAAACACGAA 1078
      ::::: ||::: |||
335 rGlyGlnHisHisTrpLysThr.....ThrGlyThr. 345
      |||
1079 TTCAATTGTTAATGTTCTTTATCCGACAGACAGAACCTGTTAT 1128
      |||
346 .....AsnSerHisIleGlySerThr..... 352
      |||
1129 CATGCTGAGGTGTTGTCACAGATTATGACCCAGACTGAATATGAGA 1178
      ||::: ||::: ||| |||||
353 ...AlaValArgLeuAlaAsnAsnGlnArgAsp...AlaAsnAsnGly 367
      ||| ||| |||
1179 AATATTTCTTTATTGACGAAAGAAAGCGAATTGATCTACTACCA 1228
      ||::: ||| ||::: ||| ||::: |||
367 nasnValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGln 382
      |||
1229 ACATCAATCAAGGTGCTGAGATATATTTCCAGAGATTTTACG 1278
      ||||| ||||| ||||| |||||
382 snIleAsnGlnIleValAlaGlyLeuPhePheLysGlyAspTyrThrVal 398
      ||||| ||||| ||||| |||||
1279 TCGCTGAAATACGAA...ACTTGGCAAGCGCGGCTTCATATCAG 1325
      ||||| ||||| ||||| |||||
399 LysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspVal 415
      ||||| ||||| ||||| |||||
1326 TGAACACAGTACCTTACTTGAAGTAAAGCGCTGCAACACGCGCC 1375
      ::::: ||::: ||||| |||||
415 aspIlyLysLysValIleTrpGlnValLysAsnProAsnGlyAspArg 432
      ||::: ||| ||||| |||||
1376 TGTCCAAATCGCAAGGACGCTGACGTTCAAGCCAAAGGGGAAAG 1425
      ||::: ||| ||||| |||||
432 euAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsn 448
      ||||| ||||| ||||| |||||
1426 CAAGCTCGATCAGCGTGGGCGAGCGTACAGTCAATTTGGATCAG 1475
      ||||| ||||| ||||| |||||
449 GlnGlyGluLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAl 465
      ||||| ||||| ||||| |||||
1476 AGACGATTAAGGCAAAAAACAGCCTTAGTGAATCGGCTTGACAG 1525
      ||||| ||||| ||||| |||||
465 aspSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValSerG 482
      ||||| ||||| ||||| |||||
1526 GCAGGGGTACGGTGCACACTGAATGCCGATTAATCACTTCAACCC 1575
      ||||| ||||| ||||| |||||
482 LysArgIlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 498
      ||||| ||||| ||||| |||||
1576 CTCATATTCGGCTTTCGGGGCGGACGTTTGATTTAAAGGGCATTC 1625
      ||||| ||||| ||||| |||||
499 LeuTyrPheGlyPheArgGlyLysArgLeuAspAlaAsnGlyAsnAsp 515
      ||||| ||||| ||||| |||||
1626 TTGCTTCACCGTATTCAAAATACGATGAAGGGCGATGATTTGCA 1675
      ||||| ||||| ||||| |||||
515 uThrPheGlnHisIleArgAsnValAspGlnValArgIleValAsnH 532
      ||||| ||||| ||||| |||||
1676 ACAATCAAGACAAAGATCCACCTTACCATTAACGCG... 1713
      ||||| ||||| ||||| |||||
532 HisAsnThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeu 548
      ||||| ||||| ||||| |||||
1713 ..... 1713

```

```

549 ThrAsnProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAsp 565
      |||||
1714 .....AAT 1717
      |||||
565 uAspAspTyrSerTyrTyrTrpArgProArgProIleProGlnGly 582
      |||||
1718 AGATATT..... 1725
      |||||
582 yAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGly 598
      |||||
1726 .....GCTACACCCGCAATAC.... 1743
      |||||
599 SerValAsnAlaProMetProGlnAsnGlyGlnThrGluAsnAspTr 615
      |||||
1744 .....ACA 1747
      |||||
615 PileuMetGlySerThrGlnGluAlaLysLysAsnAlaMetAsnH 632
      |||||
632 LysLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGlu 648
      |||||
1798 GATACGACCAAAACGAAAGCGGCTCAACCTTTTACCAGCCGCC 1847
      ||||| ||||| |||||
649 AsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLys 665
      ||||| ||||| ||||| |||||
1848 AGAAGACCGCACCCCTGCTGCTCCGCGGACCAATTTAAACGCA 1897
      ||||| ||||| ||||| |||||
665 rAlaGlnAsnArgPheLeuLeuThrGlyLysThrAsnLeuAsnGlyLys 682
      ||||| ||||| ||||| |||||
1898 TCACGCAAAACGCAACGCTTTTTCAGCGGACGACCAACCGCAC 1947
      ||||| ||||| ||||| |||||
682 LeuSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrPro 698
      ||||| ||||| ||||| |||||
1948 GCCTACATATTTAAACGACCACTGTGCTCCGCAAAAGAGGCC.... 1991
      ||||| ||||| ||||| |||||
699 AlaArgAspPheValAsnLysSerSerAlaArgLysAspAlaArgPhe 715
      ||||| ||||| ||||| |||||
1992 TCTCGCGGGGAATCGTGGGACACGCTGATCAACCGCATTTA 2041
      ||||| ||||| ||||| |||||
715 rLysAsnAsnGlnValValPheGlnAspArgPheAsnArgThrPhe 732
      ||||| ||||| ||||| |||||
2042 AAGCGGAAACCTTCAAAATTAAGCGGACGAGCGGTGCTTCC...CGC 2088
      ||||| ||||| ||||| |||||
732 yAlaAlaGlnIleAlaValAlaValAsnGlnSerAlaSerPheSerGlyArg 748
      ||||| ||||| ||||| |||||
2089 AATGTTCCCAAGTGAAGCGCATGTCATTGACGAATCAGCCGCAAC 2138
      ||||| ||||| ||||| |||||
749 AsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysVal 765
      ||||| ||||| ||||| |||||
2139 AGTTTGTGTCGCAACCGCATCAAGCCACACAACTGTACAGTTCG 2188
      ||||| ||||| ||||| |||||
765 LysnLeuGlyTyr....LysAsnGlyAspGlnValLysValArgSer 780
      ||||| ||||| ||||| |||||
2189 ACTGGAGGGTCTGACAAATTTGTGCAAAAAACCATTCACGACGATA 2238
      ||||| ||||| ||||| |||||
780 sPyrThrGlyTyrValThrCys..... 787
      ||||| ||||| ||||| |||||
2239 GTGATTGCTTCATTGACTAGACCGCATCAGCGGCAATGTCATCTGC 2288
      ||||| ||||| ||||| |||||
788 .....AsnThrAsp.....AsnLeuSer 793
      ||||| ||||| ||||| |||||
2289 CGATACAGCTCATTTAATCTACAGGGCTTGCACACTCAACGGCAATC 2338
      ||||| ||||| ||||| |||||
793 rAspLysAla.....LeuAsn..... 798
      ||||| ||||| ||||| |||||
2339 TTAGTGCAAATGGCGATACAGTTATACAGTACGCCACACGACCCAA 2388
      ||||| ||||| ||||| |||||
799 .....SerPheAspAlaThrGln 804
      ||||| ||||| ||||| |||||
2389 ...AAGCGCAACCTTAGCCCTGTTGGCAATGCCAGCAACATTTATCA 2435
      ||||| ||||| ||||| |||||
805 IleAsnGlyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGly 821
      ||||| ||||| ||||| |||||

```



```
2436 AGCCACATTAAACGGCACACATCGGCTTCGGCAATGCTTCAATTAAATC 2485
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
821 sAlaIaLeuTrpGlyGlnIleGlnIlyGlnIlyAsnSerIaISeI 838
2486 TAAGCGACCGCGCGGTACAAAAGCGAGTGTGACGCTTTCGGCAACGCT 2535
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
838 euAsnGlnHis..... 841
2536 AAGGCAAGGTAAAGCCATTCCGACGTCAAGGTAATGTCTCCCTAGCCGA 2585
841 ..... 841
2586 TAAGGAGTATTTCATTTTGAAGAGCGCGCTTACGCGAATAATCAGCG 2635
841 ..... 841
2636 GCGGCAAGGATACGGCATTTACACTTAAAGACAGCGAATGACGCTGCGG 2685
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 ..... 842
2686 TCAGGACGCGAATTAGCAATTAAACCTTGACAACGCGCACCATTAACCT 2735
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
848 SerAspSerGlnValHisAsnLeuSerIeuaIaAspSerHisIleHisIe 864
2736 CAATTCCCGCTATGCCACAGATGCGGCGGCGCAAAACGCGCAGTGCGA 2785
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
864 uAsnAsnHis.....SerAspIaGlnSerIa..... 873
2786 CAGATGCGCGCGCGCGCTTCGCGCGCTTCCGCGCTCCCTATTATTCC 2835
873 ..... 873
2836 GTTACACGCGCAACTTCGAGTAATCCGTTTCAACAGCTGACGGAATA 2885
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
874 ..... 874
2886 CGGCAATTGAACGCTCAGGAGACATTCGCTTATGTGCGAATCTTCG 2935
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
882 n...HisLeuSerGlyAsnGlyHisIleHisIleHisIleHisIleHisIe 898
2936 GCTACCGGACGACAAATTTGAAGAGTTCGCGAAGTTCGCGAAGCAGCTTAC 2985
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
898 yAsnLeuGlyAspIlyValIeuaIlyGlnSerIaIaSerGlyHisIle 914
2986 ACCTTGGCGGTCAACAATACCGGACAGCACTGCAAGCTTCGACAAATT 3035
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
915 GlnLeuHisValGlnAspIlyThrGlyIlePro...AsnGlnIleGly 930
3036 GACGCTAGTGAAGAAAGCAACAACCGCTGTCCGAAACCTTAATT 3085
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 uAsnLeuHisAspIaSerIeValGlnAspArgSer...ArgLeuSerIy 946
3086 TCACCTCTGCAAAAGCAACAGCTGATGCGGCGGTGCGGCTTCCCAATC 3135
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 alSerLeuAlaAsnHisValAspLeuGlyAlaLeuArgIlyThrIle 962
3136 ATCCGCAAAAGCGCGAGTTCGCGCTGCAATATCCGCTCAAGAA..... 3180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 LysThrGlnAsnGlyIleThrArgLeuIyAsnProIyAlaGlnHisnArg 979
3180 ..... 3180
979 gArgArgValIyProIaIaProSerProIaIaHisnThrAlaSerGln 996
3180 ..... 3180
996 IagIlnHisSerAlaGlnIleAlaIyProGlnAsnIleValAla 1012
3181 ..... 1012
1013 ProProSerProGlnAlaAsnGlnIaGlnIaIyAspArgGlnGlnAl 1029
3204 CAAGCAGAA..... 3213
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1029 alysAlaIagIlnValIyAspArgGlnGlnAlaGlnIaIyIyValAlaHis 1046
3213 ..... 3213
1046 IagIlnLysAlaGlnIaIyAlaIyAspArgGlnGlnAspAlaIeuaIaArgGln 1062
3213 ..... 3213
1063 GlnAlaIagIlnGlnIyArgGlnArgLeuGlnIaIaGlnIaIyArgGlnAlaIagI 1079
3214 ....GCCAAAAACAGCGGGAAGAAAGACACGCGCAAGCTTGACGCGC 3259
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1079 uIleAlaIyGlnIyAlaIagIlnIaIagIlnIaIyAlaIyAspArgAlaIag 1096
3260 TGATTGCGCGCGCGCGATGCGCGTCAAGAAAGACAGAA...AGCGTTCGC 3306
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1096 IuIleAlaGlnIlnIyAlaIaIaIaIagIlnIaIyAlaIyAspArgAlaIag 1112
3307 GAACGCGCGCGCGAG.....GCAGCGGCGGCAAAATGTCGCG 3341
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 GlnIeuaIaIaArgGlnGlnIlnIyAlaIaIyGlyAlaIaIagIlnIeuaIaI 1129
3342 CATTAATGACGCGCGGAGAGAGAAAAA.....C 3370
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1129 alysGlnIyAlaIagIlnIyArgGlyAlaIaIagIlnIealagIlnI 1146
3371 GGGTCAGCGGATTAAGACACCGCC...TTGGCAACAGCGC...GAA 3414
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1146 yAlaGlnIaIagIlnIyArgGlnIaIaIagIlnIeuaIaIyGlnIyAlaIagI 1162
3415 GCGGAACCGCGCGCGCTACACCGCGCTCCCGCGCGCGCGCGCGCGC 3464
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 GlnIyGlyArgGlnAlaIaIagIlnSerGlnProIyAspArgGlnAsnArg 1179
3465 C.....CGGAGTTTCGCGC 3478
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179 gAlaIleProProGlnIeuaSerIeSerAlaIaIaIaIaIeuaIaIePro 1196
3479 AACTG..... 3483
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1196 rGlnIleAlaArgAsnSerAsnProAspAlaSerAspIyGlnIleIlePro 1212
3483 ..... 3483
1213 LeuAspAlaIeuaGlnAspGlnAspValSerGlnSerValAspThrSerAs 1229
3483 ..... 3483
1229 pIyGlnProGlnAspAsnThrGlnIeuaHisGlnIyValGlnThrValS 1246
3484 ..... 3484
1246 eIleGlnProArgAlaIaIagIlnProArgAlaIagIlnAlaIaIagIlnPro 1262
3490 CAACCGGACGCCCAACCGCGCGCGCGATGATGACCGCTTAATGCCAATAG 3539
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1263 GlnIaIagIlnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 1279
3540 CGGTTGAGTGAATTTCCGCGCGCTCAAGACGCTTTCGCGC...GTAC 3586
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 rAlaIeuaSerAspAlaIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 1296
3587 AGGACGAATTAGACCGCGGTATTTCGCGAAGACCGCGC..... 3624
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1296 IyAlaIeuaSerIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 1312
3625 ....AACGCGTTTGGACAGCGCATTCGCGGACACCAAAACCTTACCGTTC 3671
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1313 LysAsnSerValIlyIleIleIleIleIleIleIleIleIleIleIleIleIleI 1339
3672 GCAGATTTTCGCGCGCTTACCGCAACAAACGACCTGCGCAATTCGCTA 3721
```



```

665 GTCCGATATCTGGCTGCTGGCGAATACCTTGCACAAAATGATCA 714
      ::::::::::: ::::::::::: :::::::::::
237 sPlatYthrThrGlyIleAlaGlyThrProTyrLysValAsnIseIu 253
      ::::::::::: ::::::::::: :::::::::::
715 GGGGGGGCAGCAGTCACTAGTAGTGAATAAATTAACATAGC...CC 761
      ::::::::::: ::::::::::: :::::::::::
254 AsnAsnGlyLeuIleGlyPheGlyAsnSerLysGluIleHisSerAsp 270
      ::::::::::: ::::::::::: :::::::::::
762 AATAGGTTTTTACCAACAGCA...GGCCATTTGGCG 796
      ::::::::::: ::::::::::: :::::::::::
270 oLysGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGly 287
      ::::::::::: ::::::::::: :::::::::::
797 ACAGTGGCTCACCAATGTTATCTATGATGCCAAACCAAAAGCTGTTA 846
      ::::::::::: ::::::::::: :::::::::::
287 sPserGlySerProLeuPheValTyrAspArgGlyLysGlyLysThrLeu 303
      ::::::::::: ::::::::::: :::::::::::
847 AATTAAATGGGCTATGCAACGGCAGCCCTATATAGAAAACCAATGG 896
      ::::::::::: ::::::::::: :::::::::::
304 PheLeuGlySerTyrAspPheThrAlaGlyTyrAsnLysLysSer.... 318
      ::::::::::: ::::::::::: :::::::::::
897 CTTCACGCTGCTCGTAAAGATTGG...TTCTATGATGAATCTTGCTG 943
      ::::::::::: ::::::::::: :::::::::::
319 .....TrrGlnGluTrrPasnIleTyrLysSerGlnPheThrL 331
      ::::::::::: ::::::::::: :::::::::::
944 GAGATACCCANTCACTATCTACAGACCGCAAAATGGGAATACTCT 993
      ::::::::::: ::::::::::: :::::::::::
331 ysAspVal..... 333
      ::::::::::: ::::::::::: :::::::::::
994 TTATACGACGATATATGACGACAGAAATCAATGCCAAACATGACA 1043
      ::::::::::: ::::::::::: :::::::::::
334 LeuAsnLysAspSerAlaGlySer...LeuIleGlySerLysThrAspTy 349
      ::::::::::: ::::::::::: :::::::::::
1044 CAATCTCTGCTATATAGATTAAACAGCAACCGTT.....C 1081
      ::::::::::: ::::::::::: :::::::::::
349 rSerTrpSerSerAsnGly...LysThrSerThrIleThrGlyGlyLul 365
      ::::::::::: ::::::::::: :::::::::::
1082 AATGTTAATGTTCTTTATCCGACAGACAGACAGAACCTGTTTATCAT 1131
      ::::::::::: ::::::::::: :::::::::::
365 ysSerLeuAsnValAspLeuAlaAspLysAspLysPro..... 378
      ::::::::::: ::::::::::: :::::::::::
1132 GCTCGAGGTGGTGTCAACAGTTATCGACCCAGCACTGAATGAATGA 1181
      ::::::::::: ::::::::::: :::::::::::
379 .....AsnHisGlyLysSe 383
      ::::::::::: ::::::::::: :::::::::::
1182 TATTTCCCTTATTGACAGAGAAAGCGAATTGATCTTACAGCAGACA 1231
      ::::::::::: ::::::::::: :::::::::::
383 rValThrPhe....GluGlySerGlyThrLeuThrLeuAsnAsnAsnI 398
      ::::::::::: ::::::::::: :::::::::::
1232 TCATTCAGAGGTGGTGGAGATTATTTCCAGAGAGATTTCACGGTC... 1278
      ::::::::::: ::::::::::: :::::::::::
398 leAspGlnGlyAlaGlyGlyLeuPheGlyGlyLysPtyrGluValLys 414
      ::::::::::: ::::::::::: :::::::::::
1279 TCGCCTGAAAAATAAGAACTGGCAAGCGCGCGCTTCATTCATGCTAGA 1328
      ::::::::::: ::::::::::: :::::::::::
415 GlyThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaG 431
      ::::::::::: ::::::::::: :::::::::::
1329 AGACAGTACCGTTACTTGGAAATAACGGCGTGGCAAAAGCAGCGCTGT 1378
      ::::::::::: ::::::::::: :::::::::::
431 uGlyLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeu 448
      ::::::::::: ::::::::::: :::::::::::
1379 CCAAAATCGCAAGAGCAGCTGCAGCTTCACCAAGGAGGAAAAACCA 1428
      ::::::::::: ::::::::::: :::::::::::
448 lAlysIleGlyLysGlyThrLeuIleValGluGlyThrGlyLysAsnLys 464
      ::::::::::: ::::::::::: :::::::::::
1429 GCGTCGATCAGCGTGGCGCAGCTACAGTCATTTGGATCAGCAGCAGCA 1478
      ::::::::::: ::::::::::: :::::::::::
465 GlySerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAs 481
      ::::::::::: ::::::::::: :::::::::::
1479 CGATAAAGCGCAAAACAGCCTTAGTGAATCGGCTGGTCAGCAGCGGA 1528
      ::::::::::: ::::::::::: :::::::::::
481 nGlySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyA 497
      ::::::::::: ::::::::::: :::::::::::

1529 GGGGTCAGGTGCACTGAATGCCGATTAATCACTTCAACCCGACAAATC 1578
      ::::::::::: ::::::::::: :::::::::::
497 rGserThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIle 513
      ::::::::::: ::::::::::: :::::::::::
1579 TATTTGCGCTTCCGGCGGAGCTTGGATTAAACGGCGATTCGGCTTTC 1628
      ::::::::::: ::::::::::: :::::::::::
514 TyrPheGlyPheArgGlyGlyValArgLeuAsnLysAsnGlyAsnSerLeuThr 530
      ::::::::::: ::::::::::: :::::::::::
1629 GTTCCACCGTATTCAAAATACCGATGAAGGGCGCATGATGTGCACACA 1678
      ::::::::::: ::::::::::: :::::::::::
530 rPheAsnHisIleArgAsnIleAspAspGlyAlaValArgLeuValAsnHisA 547
      ::::::::::: ::::::::::: :::::::::::
1679 ATCAGACCAAAAGATCCACCGTTACATTACAGC..... 1713
      ::::::::::: ::::::::::: :::::::::::
564 AspProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAs 580
      ::::::::::: ::::::::::: :::::::::::
1713 ..... 1713
      ::::::::::: ::::::::::: :::::::::::
580 nProTyrAlaPheArgArgIleLysAspGlyGlyLeuTyrLeuAsnL 597
      ::::::::::: ::::::::::: :::::::::::
1713 ..... 1713
      ::::::::::: ::::::::::: :::::::::::
597 euGluAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSer 613
      ::::::::::: ::::::::::: :::::::::::
1713 ..... 1713
      ::::::::::: ::::::::::: :::::::::::
614 GluLeuProLysAsnSerGlyLysSerAsnGluAsnTrpLeuTyrMetG 630
      ::::::::::: ::::::::::: :::::::::::
1714 ....AATAAGATATTGCTACACCGGCAATTAACACAGCTTGAGATAGA 1759
      ::::::::::: ::::::::::: :::::::::::
630 yLysThrSerAspLysValAlaLysArgAsnValMetAsnHisIleAsnAsnG 647
      ::::::::::: ::::::::::: :::::::::::
1760 AAAAAAGAAATGCTTACACAGCGTTGGTGGCGGAGAAAGATACGACCAAA 1809
      ::::::::::: ::::::::::: :::::::::::
647 lValArgMetAsnGlyPheAsnGlyTyrGlyPheGlyGluGluGlu...GlyLys 662
      ::::::::::: ::::::::::: :::::::::::
1810 ACGAAGCGGCGGCTCAACCTTGTTTACAGACCGCGCGCAGAAAGCCGCA 1859
      ::::::::::: ::::::::::: :::::::::::
663 AsnAsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnTr 679
      ::::::::::: ::::::::::: :::::::::::
1860 CCTGCTGTTTCCGGGAGAACAAATTAACGGCAATCAGCAACCAACA 1909
      ::::::::::: ::::::::::: :::::::::::
679 gPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluL 696
      ::::::::::: ::::::::::: :::::::::::
1910 ACGGCAAACTGTTTTCAGCGCGCAGCAACACCGCAGCTTACANTCAT 1959
      ::::::::::: ::::::::::: :::::::::::
696 ysGlyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIle 712
      ::::::::::: ::::::::::: :::::::::::
1960 TTAACGACCATTTGTCGCAAAAAGAGGC.....ATTCTCGCGGGGA 2003
      ::::::::::: ::::::::::: :::::::::::
713 lAlaGlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnG 729
      ::::::::::: ::::::::::: :::::::::::
2004 AATGCTGTGGAGACAGCAGCTGATCAACCGCACTTAAGGGAAGAAAT 2053
      ::::::::::: ::::::::::: :::::::::::
729 uValValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrHis 746
      ::::::::::: ::::::::::: :::::::::::
2054 TCCAAATTAAGCGCGACAGCGGCTTTC...CGCAATGTGCCAA 2100
      ::::::::::: ::::::::::: :::::::::::
746 etAsnValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsn 762
      ::::::::::: ::::::::::: :::::::::::
2101 GTGAAGGCGATTGGCATTTGAGCAATCAAGCCCAAGCAGTTTGGTGT 2150
      ::::::::::: ::::::::::: :::::::::::
763 lIleThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTy 779
      ::::::::::: ::::::::::: :::::::::::
2151 CGACCGCATCAAAAGCCACACAATCTGTACAGCTTCGAGTGCAGCGGCTC 2200
      ::::::::::: ::::::::::: :::::::::::
779 r.....LysThrGlyAspThrValCysValAlaArgSerLysPtyrThrGlyT 794
      ::::::::::: ::::::::::: :::::::::::
2201 TGACAAATTTGTGCAAAAAAACCATTTACCGAGCATAAAGTATGCTTCA 2250
      ::::::::::: ::::::::::: :::::::::::

```

```

794 yValThrThrAspLysLeuSer..AspLysAlaLeuSer 809
      :::::|||||:      ::::: |||||:      |||
2251 TTGACTAAGCAGCATCGAGCAATGCTGATCTTGCGATCATCGCTCA 2300
      ::::: |||:      |||||:      ||::: |||:
810 PheAsnProThrAsnLeuArgLysValAsnLeuThrGluSerAlaAs 826
      ::::: |||:      |||||:      ||::: |||:
2301 TTTAAATCTCAGAGGGCTTGCCACACTCAGCGCAATCTTACTGCAATG 2350
      ::::: |||:      |||||:      ||::: |||:
826 nPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArg 842
      ::::: |||:      |||||:      ||::: |||:
2351 GCGATACAGGTTATACAGTACGACCAACCCCAAAACGCAACCTT 2400
      ::::: |||:      |||||:      ||::: |||:
842 LysAsnSerGlnValArgLeuThrGluAsn..... 851
      ::::: |||:      |||||:      ||::: |||:
2401 AGCCTGTGGGCAATGCCAGCAACATTTAATCAAGCCACATTAAACGG 2450
      ::::: |||:      |||||:      ||::: |||:
851 ..... 851
2451 CAACACATCGGCTTGGGCAATGCTTCAATTAATCAAGCAACGCGCG 2500
      ::::: |||:      |||||:      ||::: |||:
851 ..... 851
2501 TACAAAACGCGAGTCTGACGCTTCCGGCAACGCTAAGCAACGTAAAGC 2550
      ::::: |||:      |||||:      ||::: |||:
851 ..... 851
2551 CATTCGCGACTCAACGGTAATGCTCTCCCTAGCCGTAAGCGAGTATCCA 2600
      ::::: |||:      |||||:      ||::: |||:
851 ..... 851
2601 TTTTGAAGACGCGCTTACGGCAACAATCAGCGGCGCAAGATACGG 2650
      ::::: |||:      |||||:      ||::: |||:
851 ..... 851
2651 CATTACACTTAAAGACAGCAATGAGCGCTGCCGACGACGCAATTA 2700
      ::::: |||:      |||||:      ||::: |||:
852 .....SerHisTrpHisLeuThrGlyAsnSerAspVal 862
      ::::: |||:      |||||:      ||::: |||:
2701 GGCATTTAAACCTTGACAAAGCCACCATTAACATTCGCCCTATCG 2750
      ::::: |||:      |||||:      ||::: |||:
863 HisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAsp 879
      ::::: |||:      |||||:      ||::: |||:
2751 CCAGGATCGCGCAGGGCGCAACCGCGAGTGGCAGATGCCGCGCGC 2800
      ::::: |||:      |||||:      ||::: |||:
879 nSerAsn..... 881
2801 GCCGTGCGCGCTTCCGCGCTTCCCTAATTCGTTACACGCGCAACT 2850
      ::::: |||:      |||||:      ||::: |||:
882 .....AsnValThr..... 884
      ::::: |||:      |||||:      ||::: |||:
2851 TCGGTAGAATCCGTTCAACACGCTGACGTAACGGCAATTTGACGG 2900
      ::::: |||:      |||||:      ||::: |||:
885 .....LysTyrAsnThrLeuThrValAsn...SerLeuSerG1 896
      ::::: |||:      |||||:      ||::: |||:
2901 TCAGGAACATTCGCTTATGCGAACTCTTGCGCTACGCGACGCA 2950
      ::::: |||:      |||||:      ||::: |||:
896 YAsnGlySerPheTyrTrpLeuThrAspLeuSerAsnLysGlnLysPl 913
      ::::: |||:      |||||:      ||::: |||:
2951 AATTGAACCTGGCGAAAATTCCGAAGGCACTTAACCTTGCGCGTCAAC 3000
      ::::: |||:      |||||:      ||::: |||:
913 yValValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAla 929
      ::::: |||:      |||||:      ||::: |||:
3001 AATACCGGCAAGCAACCTGCAACGCTGCAACATTTGAGGTAGGAGG 3050
      ::::: |||:      |||||:      ||::: |||:
930 AspLysThrGlyGluPro...AsnHisAsnGlnLeuThrLeuPheAspAl 945
      ::::: |||:      |||||:      ||::: |||:
3051 AAAAGACAACAACCGCTGTCCGAAAACCTTAATTTCAACCTGCAAAACG 3100
      ::::: |||:      |||||:      ||::: |||:
945 a.....SerLysAlaGlnArgAspHisLeuAsnValSerLeuValGly 960
      ::::: |||:      |||||:      ||::: |||:
3101 AACACGTGATGCCGCGCGCTGCGCTTACCAACTCATTCGCAAAAGACGGC 3150
      ::::: |||:      |||||:      ||::: |||:

```

```

960 snThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAsnGly 976
      ::::: |||:      |||||:      ||::: |||:
3151 GAGTCCGCGCTGCATTAATCCGGTCAAAAGAACAA..... 3183
      ::::: |||:      |||||:      ||::: |||:
977 ArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAs 993
      ::::: |||:      |||||:      ||::: |||:
3183 ..... 3183
993 pThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerV 1010
      ::::: |||:      |||||:      ||::: |||:
3184 .....GAGCTTCCGACAAACTCGGCAAG..... 3207
      ::::: |||:      |||||:      ||::: |||:
1010 alProSerAsnAsnGlnGluIleAlaArgValAspGluAlaProValPro 1026
      ::::: |||:      |||||:      ||::: |||:
3208 .....GCGAGAC 3215
      ::::: |||:      |||||:      ||::: |||:
1027 ProProAlaProAlaThrProSerGluThrThrGluThrValAlaGluAs 1043
      ::::: |||:      |||||:      ||::: |||:
3216 CAAAAACAGCGGAAAAA.....GACACGCGCAAAAGCTTGACG 3256
      ::::: |||:      |||||:      ||::: |||:
1043 nSerLysGlnGluSerLysThrValGluLysAsnGlnGluAspAlaThrG 1060
      ::::: |||:      |||||:      ||::: |||:
3257 CGCTGATTCGCGCGCGCGGCGGCGATGCCGCGAA..... 3288
      ::::: |||:      |||||:      ||::: |||:
1060 LurThrThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAsnVal 1076
      ::::: |||:      |||||:      ||::: |||:
3289 AAGACAGAAGCGTTGCCGAACCGCGCGGAGCGAGCGGGAATGT 3338
      ::::: |||:      |||||:      ||::: |||:
1077 LysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrLy 1093
      ::::: |||:      |||||:      ||::: |||:
3339 CGGCATTATGCAGCGCGAG.....GAG 3361
      ::::: |||:      |||||:      ||::: |||:
1093 sGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLysGluG 1110
      ::::: |||:      |||||:      ||::: |||:
3362 AGAAAAAGCGGTGCGAGCGGATAAGACACCGCTTGCGCAAA..... 3405
      ::::: |||:      |||||:      ||::: |||:
1110 LulysAlaLysValGluThrGluLysThrGlnGluValProLysValThr 1126
      ::::: |||:      |||||:      ||::: |||:
3406 .....CAGCGCGAGAC 3416
      ::::: |||:      |||||:      ||::: |||:
1127 SerGlnValSerProLysGlnGluGlnSerGluThrValGlnProGlnAl 1143
      ::::: |||:      |||||:      ||::: |||:
3417 GGA..... 3420
      ::::: |||:      |||||:      ||::: |||:
1143 aGluProAlaArgGluAsnAspProThrValAsnIleLysGluProGlns 1160
      ::::: |||:      |||||:      ||::: |||:
3420 ..... 3420
1160 eGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
      ::::: |||:      |||||:      ||::: |||:
3420 ..... 3420
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
      ::::: |||:      |||||:      ||::: |||:
3421 .....ACCAGCGCGCTACACCGCTTCC 3445
      ::::: |||:      |||||:      ||::: |||:
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
      ::::: |||:      |||||:      ||::: |||:
3446 CC.....CGCGCGCGCGC 3459
      ::::: |||:      |||||:      ||::: |||:
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
      ::::: |||:      |||||:      ||::: |||:
3460 GCCCGCGCGGATTTGCCG...CACTGCAACCCCAACCGCACCCCAACC 3506
      ::::: |||:      |||||:      ||::: |||:
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
      ::::: |||:      |||||:      ||::: |||:
3507 GCGCGCG.....GACGTGATCAGCCGCTTATGCAATA 3538
      ::::: |||:      |||||:      ||::: |||:
1242 nAspArgSerThrValAlaLeuLysAspLeuThrSerThrAsnThrAsnA 1259
      ::::: |||:      |||||:      ||::: |||:
3539 GCGTTTGAAGTGAATTTCCGCC.....ACGCTCAACAGC 3573
      ::::: |||:      |||||:      ||::: |||:
1259 lValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
      ::::: |||:      |||||:      ||::: |||:

```

```

3574 GTTTCGGCCGTAAGAGCAATTAAGACGGTA...TTGCCGACAGACCG 3620
|||||.....:
1276 GLYSLALALVALSERGINHISIISESGINLEUIMETASNSNGI 1292
|||||.....:
3621 CCGCAGACCGCGTTTGACACAGCGGCATCCGGACACCAACTACCGTT 3670
|||||.....:
1292 YGLNTYASINVALTRYPALSERANTHRISERMETASINLYASNTYSEIS 1309
|||||.....:
3671 CGCAAGANTTCCGGCCTACCGCCACAAACCAACCTGCGCCCAATCGGT 3720
|||||.....:
1309 ERSEGINTRYATGARGPHESESERLYSERLHRGINHTRGINLEUGLY 1325
|||||.....:
3721 ATGCAGAAAAACCTCGCAGACGG...CGCGTCGCATCCTGTTTTCGA 3767
|||||.....:
1326 TRPSPGINHTRILESERASINASINVALGINLEUGLYGLVALPHETHRY 1342
|||||.....:
3768 CAACCGACCGGAAACACCTTCAGACAGCGGCATCGCAACTCGGACGGC 3817
|||||.....:
1342 RVALARGASINERASINASINPHEASPLYSALATHRISERLYSASNTHT...L 1358
|||||.....:
3818 TTGCCACAGCGCGCGCTTTTCGGCATACGGCATTCGAC...AGGTTCTAC 3864
|||||.....:
1358 EULAGINVALASINPHEIYRISERYTTRIALASPSINSTRIPYR 1374
|||||.....:
3865 ATCGGCATTCAGCGCGCGCGGCTTTTACAGCGCGCGCTTCAGACGG 3914
|||||.....:
1375 LEUGLYLEASPLEUGLYTYRGLYLSPHGINSERLYSLEUGLINTHAS 1391
|||||.....:
3915 CATCGAGGCAAAATCCGGCGCGCGCTGCTGCATACGCGCATTCAGCAC 3964
|||||.....:
1391 NHIASINLALYSPHEALARGHISHTHALAGINPHEGLYEUTHRALAG 1408
|||||.....:
3965 GATACCGCGCGGTTTCGGCGGATTCGGCATCGCAACCGCATTCGGCGCA 4014
|||||.....:
1408 LYSALAPHEASINLEUGLYASINPHEGLYIETHTPROILEALIGLYVAL 1424
|||||.....:
4015 ACCGCGTATTCGTCCAAAACCGGATTCACGCTACGAAAACGTCATAT 4064
|||||.....:
1425 ARGTYSETRYLEUSERASINLALASPHLEALALEUASPGINALARGI 1441
|||||.....:
4065 CCGCACCGCGCGCGCTTCGATTCACCGCATCCGCGCGCGCATTAAGCGAG 4114
|||||.....:
1441 ELYSVALASINPROILESERVALYSLTHALAPHEALAGINVALASPLEUS 1458
|||||.....:
4115 ATTATTCATTCAAACCGCGCGCACACATTCCTACCGCTTATTTAGC 4164
|||||.....:
1458 ERTYTRHTYRHSLEUGLYGU...PHESEVALTHTRPROILEUSER 1473
|||||.....:
4165 CTGTCTATATCCGATCCGCTTCGGCAAAAGTCCGACACCGCGCATAC 4214
|||||.....:
1474 ALAARGTYR...ASPALASINGLINELYSERGLYLSILEASINVALASNGI 1489
|||||.....:
4215 CCGCGTATTCGCTCAGATTCGCGCAAAACCGGATTCGGAATGGGCG 4264
|||||.....:
1489 YTYRASPHLEALATYASINVALIGLUSNGLINGLINTYRASINLAGLYL 1506
|||||.....:
4265 TAAAGCCGGAATCAAGGTTTCACGCTGCTCCACGCGCGCGCGC 4314
|||||.....:
1506 EULYLEULYSTRHSINASINVALYLSLEUSERLEULIEGLYLYLEUTHR 1522
|||||.....:
4315 AAAGCCCGCAACTGGAAGCGCAACACAGCGCGCATCAATTAAGCTA 4364
|||||.....:
1523 LYSALALYSGINALAGLULYSLGLNLYSTHRLAOLULEULYSEUSERPH 1539
|||||.....:
4365 C 4365
1539 e 1539

```

seq_name: p1r2:B41859

seq_documentation_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenza

```

C:Species: Haemophilus influenzae
A:Variety: strain HK393
C:date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: B41859
R:Poulsen, K.; Reinholdt, J.; Killian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:title: A comparative genetic study of serologically distinct Haemophilus influenzae
A:reference number: A41859; MUID:92234949
A:Accession: B41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1545 <PDB>
A:Cross-references: GB:M87490; NID:g148908; PIDN:AAA24967.1; PID:g148909
A:Experimental source: strain HK393
A:Note: sequence extracted from NCBI backbone (NCBI:97283)
C:superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

alignment_scores:

```

Quality: 1579.50      Length: 1735
Ratio: 1.704          Gaps: 53
Percent Similarity: 53.429      Percent Identity: 26.686

```

alignment_block:

US-09-303-518d-649 x B41859 ..

Align seg 1/1 to: B41859 from: 1 to: 1545

```

64 CGCTTCGCGCTGCTTACTTACCAATATGCCCTGCTTCGCGCATTCCTCC 113
|||||.....:
5 LYSPLHEUSLEUSNPHLEALALEUTHRALATYRALALEUTHRPr 21
|||||.....:
114 CCAAGCGTGGCGGCGGACACACTTATTCGGCATCAACCAATACATTCATC 163
|||||.....:
21 OYTRHGINALALALEUVALARGASPYALASPIRYGINLIEPHEA 38
|||||.....:
164 GCGACTTTCGGAATAAAGCAAGTTTGACAGTGGGCGCAAGATATT 213
|||||.....:
38 RGSPLHEALAGLUSNLYSGLYSPHESEVALGILYALATHASINVA 54
|||||.....:
214 GAGGTTTACCAAAAAGGAGAGTGTGGCGCAATCAATGACAAAGC 263
|||||.....:
55 GLUVALARGASPLYSASINARGPROLEUDLYSINVALLEUPROASNGI 71
|||||.....:
264 C...CGATGATTCATTTTCTGTGCTGCG...CGTAACGCGTGGCGG 307
|||||.....:
71 YLEPROMETILEASPHESERVALVALASPYALASPYARGILEALAT 88
|||||.....:
308 CATTCGGGCGCATCATATATTCGTGACCGGCAACAT..... 345
|||||.....:
88 HLEUVALASINPROGINTRYVALIGLYVALYLSHSIVALSERASINGLY 104
|||||.....:
346 .....AACGCGCGTATAACAACGTGTA 368
|||||.....:
105 VALSERGLULEUNISPHGLYASINLEUDASINGLYASINMETASINSGLYAS 121
|||||.....:
369 TTTTGTGCGGAGAGAAATCCCGATCAACATCGTTTACTTATTAATAA 418
|||||.....:
121 NALALYSLAHNIS...ARGASPYALSERSEGLUGLUSINARGTYLYTR 137
|||||.....:
419 TTGTGAAACGGAATATATAT.....AAAGCAGGAGCT 450
|||||.....:
137 HVALGILYASINGLUTYRPROTHRILYLSLEUSNGLYLYSALALVALTHR 153
|||||.....:
451 .....AAAGCCATCTTATGCGCGGATTCATATATGCGCGGCTT 491
|||||.....:
154 THRGLUSPGINALAGLILYASARGARGLUSALPYLYTYRMECPROARGLE 170
|||||.....:
492 GCATTAATTTGTACAGATGCAGACCTGTGAA...ATGACGAGTTATA 538
|||||.....:
170 WASPLYSPEVALTHRGLUVALALAPROILEGLUALASERTHRASPSERS 187
|||||.....:

```

```

539 TGGATGGCGGGAATATATCATCAAAATATATACCGTGGTGGT 588
    |||
187 eethrlaglythrtyrasnasnlyspstysrProtyrPheValtyr 203
    |||
589 ATTGGGCGAGCGCAGCAATAT..... 609
    |||
204 LeuGlySerGlyThrIlePheIleTyrgluasnGlyThrArgTyrgluLe 220
    |||
610 .TGGCATCTGATGAAGATGAGCCCAATAACCGGAAGTTCATATCAT. 657
    |||
220 utrPLeuGlyLysGluGlyInlySerasp..AlaGlyGlyTyrasnL 236
    |||
658 .....ATTGCAAGTGGCTATCTTGGCTGGTGGTGGCAATACCTTT 699
    |||
236 euLysLeuValIglYasnAlaTyrtHrTyrglyLeuAlaGlyThrProtyr 252
    |||
700 GCACAAATATGATCAGGTGGTGGCAGACATCACTAGTAGT..... 741
    |||
253 GluValasnHisgluasnaspGlyLeuIleGlyPheGlyAsnSerAsnAs 269
    |||
742 .....GAAAAATTAAACATACCCCATATGCTTTT 772
    |||
269 nglutyrIleasnProlysgluIleuSerLysPro.....L 283
    |||
773 TACCAACAGAGGCTCATTTGGGAGACGTGGTACCACCAATGTTATCAT 822
    |||
283 eutHrAsnTyraIalaValLeuGlyAspSerGlySerProLeupheValTy 299
    |||
823 GATGCCCAAAAGCAAAAGTGTATTAATTAATGGGTATTCGAAAGCGGCAA 872
    |||
300 AspArgGluLysGlyLysTrPLeuPheLeuGlySerTyraspTyrtPAl 316
    |||
873 CCCCTATATAGCAAAAGCAATGGCTTCACGTGGTTCGTAAGATTGGT 922
    |||
316 aglyTyrt.....AsnLysLysSerTrp 324
    |||
923 TCTATGATGAATCTTTGCTGAGATACCATTCAGATATCTACGAAACA 972
    |||
324 lnglutPAsnIleTyrtLysProgluPheAlaGluLysIleTyrgluIn 340
    |||
973 CGTCAAAATGGG.....AATACTCTTTTAACGACGATA 1007
    |||
341 TyrtSerAlaGlySerLeuIleGlySerLysThrAspTyrtSerTrpSe 357
    |||
1008 TAAATGGCAGCAAGAAATCATATGCCAAACATGACACATCTCTGCTA 1057
    |||
357 rAsnGly..... 359
    |||
1058 ATAGATTAAACACAGAACCGTT.....CAATGTTTATATGT 1095
    |||
360 .....LysThrSerThrIleThrGlyGlyGluLysSerLeuAsnVal 373
    |||
1096 TCTTTATCCGAGACAGCAAGAACCTTTATATCATGCTGCAGGTGGTGT 1145
    |||
374 AspLeuAlaAspGlyLysAspLysPro..... 382
    |||
1146 CAACAGATTATGACCCAGACTGATATATGAGAAATATTTCTTTATGT 1195
    |||
383 .....AsnHisGlyLysSerValThrPhe... 390
    |||
1196 ACGAAGAAAGCGCAATGATCTTACCAGACATCAATCAAGAGTGT 1245
    |||
391 ..GluGlySerGlyThrLeuThrLeuAsnAsnIleAspGlyGlyAla 406
    |||
1246 GGGAGATTATTTCCAAAGAGATTTTACGGTC...TCGCTGAAATAA 1292
    |||
407 GlyGlyLeuPhePheGlyLysAspTyrgluValLysGlyThrSerAspAs 423
    |||
1293 CGAAATCTGGCAAGCGCGGCTTCATATCAGTGAAGACAGTACCGTTA 1342
    |||
423 nHrThrTrpLysGlyAlaGlyLysSerValAlaGluGlyLysThrValT 440
    |||
1343 CTTGGAAGTAAAGCGGTGGCAAGACCGGCTGTCCAAAAATCGGCAA 1392
    |||
440 hTrIryValHisasnProgluTyrtAspArgLeuAlaLysIleGlyLys 456
    |||
1393 GGCACGCTCAGCTTCACACCCAAAGGGAACCAAGGCTGATCAGCGT 1442
    |||
457 GlyThrLeuIleValIgluGlyThrGlyAspAsnLysGlySerLeuLysVa 473
    |||
1443 GGGCAGCGGTACAGTCATTTTGGATCCACAGGCGACAGCATAAAGCAAA 1492
    |||
473 IglYAspGlyThrValIleLeuLysGluInlThrAsnLysGly...G 489
    |||
1493 AACACCCCTTTAGTGAATCGCTTGGTGGTGGGAGGAGTACGGTGCAA 1542
    |||
489 lHnIsAlaPheAlaSerValIglYIleValSerGlyArgSerThrLeuVal 505
    |||
1543 CTGAATGCCATATATCATGTTCAACCCGCAACACTATTTTCGCTTTCG 1592
    |||
506 LeuAsnAspAspLysGluValAspProAsnSerIleTyrtPheGlyPheAr 522
    |||
522 gGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleA 539
    |||
1643 AAATACCGATGAAGGCGGATGATTTGCACCAACATCAAGCAAGAA 1692
    |||
539 rGAsnIleAspGluGlyAlaArgLeuValAsnHisSerThrSerLysHis 555
    |||
1693 TCCACGCTTCCATTTACAGCAATTAAGATTTGCTACACCGGCAATPA 1742
    |||
556 SerThrValThrIleThrGlyAsp...AsnLeuIleThrAspProAsnAs 571
    |||
1743 C..... 1743
    |||
571 nValSerIleTyrtTyrtValLysProLeuGluAspAsnProTyrtAlaI 588
    |||
1743 ..... 1743
    |||
588 leArgGlnIleLysTyrtGlyTyrtGlnLeuTyrtPheAsnGluGluAsnArg 604
    |||
1743 ..... 1743
    |||
605 ThrTyrtTyrtAlaLeuLysLysAspAlaSerIleArgSerGluPheProgl 621
    |||
1743 ..... 1743
    |||
621 nasnArgGlyLysSerAsnAsnSerTrPLeuTyrtMetGlyThyGluLysA 638
    |||
1744 .....AACAGCTTGATAGCAAAAGAAATTT 1770
    |||
638 laAspAlaGlnLysAsnAlaMetAsnHisIleAsnAsnGluAlaGluMetAsn 654
    |||
1771 GCCTACAAAGGTGTGGTGGGAGAAAGATACGACCAAAAGCAAGGCGG 1820
    |||
655 GlyPheAsnGlyTyrtPheGlyGluGlu...GlyLysAsnGlnLys 670
    |||
1821 GCTCAACCTGTTTAAACAGCCCGCGCAGAAAGACCGCACCTGCTGCTTT 1870
    |||
670 nLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeu 687
    |||
1871 CCGCGGAACAATTTAAAGCGCAACATCACGCAAAACAAAGCGCAACTG 1920
    |||
687 hGlyGlyTyrtAsnLeuAsnGlyAspLeuAsnValIglInGlyTyrtHrLeu 703
    |||
1921 TTTTTCAGCGGAGCAACACCGCACCGCTACAT..... 1956
    |||
704 PheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSe 720
    |||
1957 .....CATTTAAAGCAATTTGCTGCAGAAAGAGGCACTTCTCTGCG 1999
    |||
720 rSerThrLysLysAspSerHisPheSerGlu.....AsnA 732
    |||
2000 GGGAAATGCTGTGGCAACAGACTGTGATCAACCCGACATTAAGCGGAA 2049
    |||

```

```

732  snuIvalValValGluaSPsPTripIleasnArgAsnPhelysIalThr 748
2050  AACTTCAAAATTAAAGCGAGCGCGGTTC...CGCAATTGCG 2096
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
749  AsnIleasnValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValGl 765
2097  CAAAGTAAAGCGGATGGCATTTGAGCATACGCCCAAGCGATTTTG 2146
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
765  uSerIleThrSerAsnIleThrIleSerAsnAlaIlyValHisIleG 782
2147  GTGCGCAGCCGCATCAAGCCACACATCTGTACACGTTGGACGTGAGC 2196
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
782  IlyTyr.....LysAlaGlyAspThrValCysValAlaArgSerAspTyrThr 796
2197  GGTCTGCAAAATGTGTGCGAAAAAACCATTTACCGAGCATAAAGTATGCG 2246
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
797  GlyTyrValIleThrCysThrThrAspIlySer...AspIlyAlaIleuAs 812
2247  TTCAATTGACTAGACCGACATCAGCGCAATGTCGATCTTGGCGATCAGC 2296
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
812  nSerPheAsnProThrAsnIleuArgIlyAsnValAsnIleuThrIleuSerA 829
2297  CTGATTTAAATCTCACAGGCGTTGCCACATCAGCGCATCTTGTGTCA 2346
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
829  IAsnPheValIleu...GlyLysAlaAsnIleuPheGlyThrIleGlnSer 844
2347  AATGGCGATACACGTTTATACAGTACGCCACAGCGCACCCCAAAAGCGCA 2396
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
845  ArgGlyAsnSerGlnValArgLeuThrGluAsn..... 855
2397  CCTTACCTCGTGGCAATGCCCAAGCAACATTTAATCAGCCACCATTTAA 2446
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
855  ..... 855
2447  ACGGCACACATCGGCTTGGGCAATGCTTATTATCTAAGCGACAC 2496
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
855  ..... 855
2497  GCGGTACAAAGCGAGTCTGACGCTTTCGGCAAGCTAAGCAAAAGCT 2546
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
855  ..... 855
2547  AAGCCATTCGCGACATCAACGGTAATGTCTCCCTAGCCGATAGGCAGAT 2596
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
855  ..... 855
2597  TCCATTTTGAAGACGCGCTTTACCGGACAATCAGCGCGGCAAGAT 2646
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
855  ..... 855
2647  ACGGCATTACACTTAAAGACAGCGAATGACGCTGCGGTACGAGCAGGA 2696
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
856  .....SerHisThrHisLeuThrGlyAsnSerAs 865
2697  ATTAGCAATTAACTTGACACGCGCACCATACACTCAATCCGCT 2746
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
865  pValHisGlnIleuAspLeuAlaAsnGlyHisIleHisLeuAsnSerIlaA 882
2747  ATCGCCACGATGCGGACGAGGCGCAAAACCGGACGTGCGAGATGCGCG 2796
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
882  sPAsnSerAsn..... 885
2797  CGCGCGCTTGGCGCGCTTCCGCGCTTCCCTAATTATCCGTTACACCGCC 2846
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
886  .....AsnValThr..... 888
2847  AACTTCGTAGATCCGTTTCAACAGCTGACGCTAAAGCGAAATTGA 2896
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
889  .....LysTyrAsnThrIleuThrValAsn...SerIleuS 899
2897  ACGGTACGAGAACATTCGCTTATGTGGAACATCTTGGGTACCGCAGC 2946
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
899  ergIlyAsnGlySerPheTyrIleuThrAspLeuSerAsnIlyGlnGly 915

```

```

2947  GACAAATTAAGCTGGCGGAAAGTTCCGAAGGACTTACACCTGGCGGT 2996
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
916  AspTyrValValValIleThrIlySerIleAlaThrGlyAsnPheThrLeuGlnVa 932
2997  CAACTATACCGGCAACGAACTTGCAGAGCTCGAACAATTGACGTATGCG 3046
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
932  IAlaAspTyrThrGlyIleuPro...AsnHisAsnGlnIleuThrIleuPhea 948
3047  AAGGAAAGACACAAACGCGTCTCCGAAACCTTAATTTCACCTCGCAA 3096
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
948  spAla.....SerLysAlaGlnAlaArgAspHisIleuAsnValSerIleuVal 962
3097  AACGAACACGTCGATCCGCGCGGTGGCGCTTACCAACTCATCTCCGAAGA 3146
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
963  GlyAsnThrValAspLeuGlyAlaTrpIlyIlyIlySerIleuArgAsnValAs 979
3147  CGCGGAGTTCCGCTCATATCCGCTCAAGAACAA..... 3183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
979  nGlyArgTyrAspLeuTyrAsnProGluValGluIlySerArgAsnGlnThrV 996
3183  ..... 3183
996  AlaSpThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValPro 1012
3184  .....GAGCTTTCGCAACAACCTCGGCAAG..... 3207
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1013  SerValProSerAsnAsnGlnIleAlaIleAlaArgValAspGluAlaProVa 1029
3208  .....GCAG 3211
1029  IProProAlaProAlaThrProSerGluThrThrGluThrValAlaG 1046
3212  AAGCAAAACAGGCGGAAAA.....GACACGCGCAAGCCTT 3252
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1046  IAsnSerLysGlnIleuSerLysThrValGluLysAsnGlnIleuAlaSpAla 1062
3253  GACGCGCTGATTCGCGCGCGCGGCGATCGCTCGAA..... 3288
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1063  ThrGluThrThrAlaGlnAsnArgIleValAlaLysGluAlaLysSerAs 1079
3289  ...AAGACAGAAAGCTTGGCGAAGCGCGGACGAGCAGCGAGGAGAA 3334
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1079  nValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerIlySerIleuT 1096
3335  ATGTGCGCATTTATCGAGCGGAG..... 3357
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1096  hrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLys 1112
3358  GAAGAGAAAAAGGCGTCCAGCGGATTAAGACACCGCTTGGCGAAA.. 3405
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1113  GluGluLysAlaIlyValGluThrGluLysThrGlnGluValProLysVa 1129
3406  .....CAGCGC 3412
1129  IThrSerGlnValSerProLysGlnGluIleuSerGluThrValGlnProG 1146
3413  AAGCGGAA..... 3420
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1146  IAlaGluProAlaArgIleuAsnAspProThrValAsnIleLysGluPro 1162
3420  ..... 3420
1163  GlnSerGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluTh 1179
3420  ..... 3420
1179  rSerSerAsnValGluGlnProValThrGlnSerThrThrValAsnThrG 1196
3421  .....ACCGCGCGCGGTACACCGCC 3441
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1196  IlyAsnSerValValGluAsnProGluAsnThrThrProAlaThrThrGln 1212

```


[illegible]

```

1776 ..... 1776
617 eMetGlyTyrThnGlnGluAlaArgLysAsnAlaMetAsnHisLysA 634
1777 ..... AACGTTGGTTGGCGAGAAAGATACG 1803
634 snAsnArgArgIleGlyAspPheGlyGlyPhePheAspIleGluAsnGly 650
1804 ACCAAACAGACGCGCGCTCAACCTTGTTCACAGCCCGCGCAGAGA 1853
651 LysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaG 667
1854 CGCGACCCCTGCTCTTCCGGGAGAACAAATTAAACGGCAACATCAGC 1903
667 nLysArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysLeuSerV 684
1904 AACCAACGGCGAAGCTGTTTTCAGCGGAGACCAACACCGCGCTTAC 1953
684 aLThnGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAlaArg 700
1954 AATCATTTAAAGCAGCATTTGGTCGCAAAAGAGGC.....ATTCTTCG 1997
701 AspPheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAs 717
1998 CGGGGAATGCTGTGGGACAGACAGCTGATCAACCGCACATTAAAGCG 2047
717 nAsnGluValValPheGluAspSprIleAsnAlaGlyThrPheLysAlaA 734
2048 AAAACTCCAAATTAAAGCGGACAGCGGCTGCTTC.....CGCATGTT 2094
734 IaGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2095 GCCAAATGAAAGCGCATTTGGCATTTGAGCAATCAGCCCAACCATTT 2144
751 SerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLe 767
2145 TGCTGTGCGACCGCATCAAGCCACACATCTGTACAGCTTGGAGCTGA 2194
767 uGlyTyr.....LysAsnGlyAspGluValLysValArgSerAspLys 782
2195 CGGGCTGACAAATTGTGTGCAAAAAACCATTCGCCGAGCATAAAGAT 2244
782 hArgIlyValThrCysAsnThrGlyAsnLeuSer...AspLysAlaLeu 797
2245 GCTTATGACTAAGACGACATCAGCGGCAATGTGATCTTCCGATCA 2294
798 AsnSerPheAspAlaThrArgIleAsnGlyAsnValAsnLeu..... 811
2295 CGCTCATTTAAATCTCACAGGCTTCCACACTCAACGCAATCTTAGT 2344
811 ..... 811
2345 CAAATGGGATACACGTTATACAGTCAGCCACACGCCCAACAAAGCG 2394
812 ..... AsnGln 813
2395 AACCTTAGCTCTGTGGCAATGCCCAACCAACATTATATCAAGCCACAT 2444
814 AsnAlaIleVal.....LeuGlyLysAlaIle 824
2445 AAACGGCAACATCGCTTGGGCAATGCTTCATTTAATCTAAGCGACG 2494
824 uTrpGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnH 841
2495 ACGCGGTACAAAAGCGAGTGTGACGTTTCCGGCAACGTTAAGCAAC 2544
841 IS..... 841
2545 GTAAGCATTCGCGACTCAACGTAATGTCTCCAGCGATTAAGCGAGT 2594
841 ..... 841
2595 ATTCCATTGTAAGCAGCGCTTACCGGACAAATCAGCGGCGCAAG 2644
841 ..... 841
2645 ATAGCGATTACCTTAAAGACAGCGAATGGAGCTGCCGTCAGCAGC 2694
842 ..... SerLysThrHisLeuThrGlyAspSer 850
2695 GAATTAGCAATTTAACCTTGACACAGCCACCACTTACACTCAATTCGCG 2744
852 GlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla 867
2745 CTATCGCCACAGTTCGGGACGGGCGCAACCGGACAGTGCACATCGCG 2794
867 a.....SerAspAlaGlnSerAla..... 873
2795 CGCGCGCGGCTTCCGCGCTTGGCCGCTTCCCTATATCCGTTACACG 2844
873 ..... 873
2845 CCAACTTCGTTAGAAATCCGTTTCAACAGCTGACGCTAAGCGCAATT 2894
874 ..... AsnLysThrHisThrIleLysIleAsn...HisLe 884
2895 GAACGTCAGGAGACATTCGCTTATATGCGGACCTTCCGCTACCGCA 2944
884 uSerGlyAsnGlyHisPheHisIlyLeuThrAspLeuAlaLysAsnLeuG 901
2945 GCGACAAATTAAGCTGCGGGAAGTTCCGAAGGACATTACCTTGGCG 2994
901 LysAspLysValLeuValLysGlnSerAlaSerGlyHisIlyGlnLeuHis 917
2995 GTCAACAAATTCGCGCAACGACCTGCAAGCTTCGCAACATTTGCGAGT 3044
918 ValGlnAsnLysThrGlyGluPro...AsnGlnGlyLysLeuAspLeuPh 933
3045 GGAAGGAAAGACAAACAACCGCTGTCCGAAACCTTAATTCACCTTGC 3094
933 eAspAlaSerSerValGlnAspArgSer...ArgLeuPheValSerLeuA 949
3095 AAAACGACACGTCGATGCGCGCGCTGCGCTTACCAACTCATTCGCA 3144
949 IaAsnHisIlyValAspLeuGlyAlaLeuArgIlyThrIleLysThrGlu 965
3145 GAGCGAGGTCCGCTGCATTAAT.....CCGCT 3173
966 AsnGlyIleThrArgLeuTyrAsnProTyrAlaGlyAsnGlyArgProVa 982
3174 CAAGAACACAGAGCTTCCGACAAACTCGCAAGCGCAAGACCAAAAC 3223
982 LlysProAlaProSerProAlaIleAsnThrAlaSerGlnAlaGlnLysA 999
3224 AGCGGGAAGAACACAGCGCAACGCTTGACGCGCTGATTGCGCGCGG 3273
999 IaThrGlnThrAspGlyAlaGln.....IleAlaLysPro 1010
3274 CGCGATGCCGTGCAAAAGACGAAAGCGTTGCCGACCGCGC..... 3315
1011 GlnAsnIleVal.....ValAlaProProSerProGlnAl 1022
3316 .....CGCGACGCAAGCGGGGGAATATGTCG 3340
1022 aAsnGlnAlaGluGluAlaLeuArgGlnGlnAlaLysAlaGluGlnAl 1039
3341 GCATTATTCAGAGCG.....GAGAA 3360
1039 ysArgGlnGlnAlaIleAlaGluAlaGlyValAlaArgGlnLysAspGlu 1055
3361 GAGAAAAACGCGTGCAGCGGAT..... 3384
1056 GluAlaLysArgLysAlaIleAlaGluIleAlaArgGlnGlnGluAlaAr 1072
3385 .....AAG 3388

```

```

1072 gTysAlaAlaGluLeuAlaAlaLysGlnLysAlaGluAlaGluAlaGlySer 1089
3389 ACACGCCCTTGCGAAGACGCGCAAGACCGCGCGCGCGTACAC 3438
1089 laArgGluLeuAlaArgGlnLysAlaGluAlaSerHisGlnAlaAsn 1105
3439 GCCCTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGCA 3480
1106 AlaLysProLysArgArgArgArgAlaAlaLeuProArgProAla 1122
3480 ..... 3480
1122 aproValPheSerLeuAspAspLysAlaLysAspAsnSerGluSer 1139
3480 ..... 3480
1139 erileGlyAsnLeuAlaArgValileProArgMetGlyArgGluLeuile 1155
3480 ..... 3480
1156 AsnAspTyrGluGluileProLeuGluGluLeuGluAspGluAlaGlu 1172
3480 ..... 3480
1172 uGluArgArgGluAlaThrGlnPheHisSerLysSerArgAsnArgAla 1189
3480 ..... 3480
1189 laileSerSerGluProSerSerAspGluAspAlaSerGluSerValSer 1205
3480 ..... 3480
1206 ThrSerAspLysHisProGlnAspAsnThrGluLeuHisGluLysValG 1222
3481 ..... 3481
1222 uThrAlaGlyLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAla 1239
3506 CGACGCGCATCGATCAGCCGTTATGCCAATAGCGGTTTGAGTAATTT 3555
1239 laGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAspAla 1255
3556 TCCGCGACCGTCAACAGCGT ..... TCCGCGGTCA 3587
1256 MetaLaserThrGlnSerileLeuAspThrGlyAlaTyrLeuThrAr 1272
3588 GAGCAATTAAGACCGCGTATTTGCCGAAGACCGCGCGCGTTCGA 3637
1272 gHisileAlaGlnLysSerArgAlaAspAlaGluLysAsnSerValTrpM 1289
3638 CAAGCGCATCGCGGACACCAACACTACCGTTCGCAAGATTTCCGCC 3687
1289 eSerAsnThrGlyTyrGlyArgAspTyrAlaSerAlaGlnTyrArgArg 1305
3688 TACCCCAACAACACCGACCTCGCAATCGTATGAGAAACCTTCGG 3737
1306 PheSerSerLysArgThrGlnThrGlnileGlyLeaSpArgSerLeu 1322
3738 CAGCGGG...GCGCGGCATCGCTTTTCGCAACACCGCAACCAACA 3784
1322 rGluAsnMetGlnileGlyGlyValLeuThrTyrSerAspSerGlnHis 1339
3785 CCTTCGACGACGCGATCGCAACTCGCGACGCTTCGCCCGCGCGCTT 3834
1339 hrPheAspGlnAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeu 1354
3835 TTCGGGCATATACGGCATC...GACAGTTTCAATCGCATCGACGCGG 3881
1355 TyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyAspIle 1371
3882 CGCGGGTTTTCGACGCGACGCTTCGACGCGCATCGAGGAAATTC 3931
1371 YAlaGlySerLeuArgSerArgLeuGlnThrGlnLysAlaAsnPheA 1388
3932 GCCCGCGCTGTCATTAACGCATTCAGGACAGATACCGCGCGGTTTC 3981
1388 snArgThrSerileGlnThrGlyLeuThrGlnLysAsnThrLeuLysile 1404
3982 GCGGATTCGCGATCGAACCACATCGCGCAACGCGCTATTTCCGCA 4031
1405 AsnGlnPheGluileValProSerAlaGlyileArgTyrSerArgLeu 1421
4032 AAAACGCGATTCAGCTACGAAACGTCATATCGCACCGCGCGCTTG 4081
1421 rSerAlaAspTyrLysLeuLysAspSerValLysValSerMetAla 1438
4082 CATTCACCGCTACCGCGCGGATTAAGCAGATTATTCATTAACCG 4131
1438 laValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLys 1453
4132 GCGCAACACATTCGATCAGCGCTTATTCAGCGCTGCTCATACGATGC 4181
1454 ValGlyLysLeuThrValLysProLeuLeuSerAlaAlaTyrPhe...Al 1469
4182 CGCTTCGCGCAAGTCCGACACGCGCTCAATACCGCGCGTATTCGCG 4231
1469 aaSnTyrGlyLysGlyValAlaValGlyLysSerPheAlaTyrL 1486
4232 ATTTCGCAAAACCGCGCATGCGGAAAGCGGCTAAACCGCAATCAAA 4281
1486 yAlaAspAsnGlnGlnGlnTyrSerAlaGlyValAlaLeuLeuTyrArg 1502
4282 GGTTCACGCTGTCCTCCACGCGTCCGCGCAAGACCGCGCATCTGA 4331
1503 AsnValThrLeuAsnValAsnLysSerileThrLysGlyLysGlnLeu 1519
4332 AGCGCAACACAGCGCGCATCAATTAAGCTACCGCTCG 4371
1519 uLysGlnLysSerGlyGlnLeuLysileGlnLeuArgPhe 1532
seq_name: p1r2:H64106
seq_documentation_block:
  IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (str
  N:Alterate names: Immunoglobulin A1 proteinase type 1
  C:Species: Haemophilus influenzae
  A:Variety: strain Rd KW20
  C:date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
  R:Fileschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
  ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidmar
  , D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Geoghagen, N.S.M.
  Science 269, 496-512, 1995
  A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
  A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
  A:Reference number: A64000; MUID:95350630
  A:Accession: H64106
  A:Status: nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-1694 <TIGR>
  A:Cross-references: GB:U32779; GB:I42023; MID:91574009; PIDN:AAC22651.1; PID:91574019
  A:Experimental source: strain Rd KW20
  R:Grundy, F.J.; Platt, A.G.; Wright, A.
  Infect. Immun. 58, 320-331, 1990
  A:Title: Localization of the cleavage site specificity determinant of Haemophilus inf
  A:Reference number: A41500; MUID:90129281
  A:Molecule type: DNA
  A:Residues: 1-377 <GRD>
  A:Cross-references: GB:X59800
  A:Experimental source: strain Rd KW20
  A:Note: The authors translated the codon TGG for residue 319 as Thr
  C:Function:
  A:Description: this proteinase is classified as type 1 because it cleaves at a prolin
  C:Superfamily: IGA-specific metalloendopeptidase
  C:Keywords: hydrolase, metalloproteinase

```

```
alignment_scores:
```

Quality:	1568.00	Length:	1864
Ratio:	1.673	Gaps:	50
Percent Similarity:	50.268	Percent Identity:	25.161

```
alignment_block:
```

US-09-303-518D-649 x H64106 . .

Align seg 1/1 to: H64106 from: 1 to: 1694

[illegible][illegible]

[illegible]

```

3171 ..... 3171
1000 rThrasnIeThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 ..... 3171
1017 rSerAsnAsnGlnGluIleAlaArgValGluThrProValProProPro 1033
3171 ..... 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 ..... 3171
1050 sGlnGlnSerLyThrValGluLysAsnGlnGlnAspAlaThrGluThrT 1067
3171 ..... 3171
1067 hTrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 ..... 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGlnGluThr 1100
3172 ..... GTCAAAGA... 3180
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluLysAla 1117
3181 ..... CAAAG 3186
1117 LaLysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSerGlu 1133
3187 CTTTCGCAAACTCGGCAAG... 3207
1134 ThrSerProLysGlnAlaLysProAlaProLysGluValSerThraspTh 1150
3207 ..... 3207
1150 rLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSerThrT 1167
3208 ..... GCAGAGCC 3216
1167 hTrValAlaAlaGluAlaThrSerProAsnSerLyProAlaGluGlu 1183
3217 AAAAAACAGCGCGAAAGACACAGCGCAAGCGCTGACGCGGATTCG 3266
1184 ThrGlnProSerGluLysThrasnAlaGluProValThrProValLysE 1200
3267 G.....G 3268
1200 rLysAsnGlnThrGlnAsnThrThraspGlnProThrGluArgGluLysT 1217
3269 CCGGCGCGGATCGCGTGAAGAAAGACAGAAAGCGTTCGCGAAGCGCGCG 3318
1217 hTrAlaLysValGluThrGlnLysThrGlnGluProGlnValAlaSer 1233
3319 CAGGAGGC..... 3327
1234 GlnAlaSerProLysGlnGlnSerGluThrValGlnProGlnAlaVala 1250
3328 .....GGGAAATGTC...GGCATTCACAGCGCGAGAGAGAGAAA 3367
1250 lLeuGlnSerGluAsnValProThrValAsnAsnAlaGlnGluValGlnA 1267
3368 AACGGGTCAGCGGATTAAGAC..... 3390
1267 lAglnLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnProAla 1283
3391 .....ACCGCTTGCGCAACAGCG 3410
1284 ProGluAsnSerLleAsnThrGlySerAlaThrAlaIleThrGluThrAl 1300
3411 CGAAGCGAAACCGCGCGCTAC..... 3435
1300 aGluLysSerAspLysProGlnThrGluThrAlaAlaSerThrGluAspA 1317
3435 ..... 3435
1317 lAsSerGlnIleLysAlaAsnThrValAlaAlaAspAsnSerValAlaAsnAsn 1333
3436 .....ACCGCTTCGCCCGCGCGCGCGCGCGCGCGCGGATTCGCGCA 3479
1334 SerGluSerSerAspProLysSerArgArg...ArgArgSerLleSerG 1349
3480 ACTGCA..... 3486
1349 nProGlnGluThrSerAlaGluGluThrThrAlaAlaSerThrAspGluT 1366
3487 .....CCCGACCGCGACCGCAACGCGAG 3510
1366 hTrThrLleAlaAspAsnSerLyArgSerLyProAsnAlaArgSerArg 1382
3510 ..... 3510
1383 ArgSerValArgSerGluProThrValThrAsnGlySerAspArgSerTh 1399
3511 .....CGGACCTGATCAGCCGTTATGCCAATAGGCGTTGAGTG 3550
1399 rValAlaLeuAlaArgAspLeuThrSerThrAsnThrAsnAlaValIleSerA 1416
3551 AATTTCGCGC.....ACGTCACACGCGTTTCGCGGTA 3585
1416 sPAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLysAlaValA 1432
3586 CAGGACGAATTAGACCGGTA...TTTGCGCAAGACCGCGCGCGCGCT 3632
1433 SerGlnIleIleSerGlnLeuGluMetAsnAsnGlnGluGlnTyLysAsnVa 1449
3633 TTGACACAGCGCGCATCGCGGACACCAACACACGCTTCGCAAGATTTC 3682
1449 lTrValSerAsnThrSerMetAsnGluAsnTyLysSerSerGlnTyL 1466
3683 GCGCCTACCGCGCAACCAACGCACTCGCGCAATCGTATCGAGAAAC 3722
1466 rGArgPheSerSerLySerThrGlnThrGlnLeuGluTyTrAspAlnThr 1482
3733 CTCGCGACGCGG...CGGTCGCGCATCGCTTTTCGCAACCGGAGCA 3779
1483 lIleSerAsnAsnValGlnLeuGluGlyAlaPheThrTyValAlaGlnSe 1499
3780 AACAACCTTCGACGACGCGATCGGCACTCGGACGCGCTTCGCGCGCG 3829
1499 rAsnAsnPheAspLysAlaSerSerLyAsnThr...LeuAlaGlnValA 1515
3830 CCGTTTCGCGCAATACGCGATCGAC...AGTTCATACATCGCGATCGAC 3876
1515 snPheTySerLyTyTyAlaAspAsnIleTyTyTyLeuGluIleAsp 1531
3877 GCGGCGCGCGGTTTACGACGCGCAAGCTTCGACGCGCGAGGAGCA 3926
1532 LeuGluTyGluTyLysPheGlnSerAsnLeuTyTyThrAsnIleAsnAla 1548
3927 AATCGCGCGCGCGCTGATTCAGCATTCAGTACGACATCGCGCGCG 3976
1548 sPheAlaArgIleThrAlaGlnPheGluLeuThrAlaGluLysAlaPheA 1565
3977 GTTTCGCGCGATTCGCGATCGCAACGCGCAATCGCGCGCGCGATTC 4026
1565 snLeuGlnAsnPheGluTyIleThrProIleValGluAlaArgTySerTy 1581
4027 GTCGCAAAAGCGATTCAGCGCAAGATTCAGATTCGCGCGCGCG 4076
1582 LeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLysValAsnPr 1598
4077 CCGTCGATTCACCGGCTACCGCGCGCGCATTAAGGACGATTCGATTCGA 4126

```

```

1598 oIleSeVallysthrAlaPheAlaGlnValAspLeuSerTyrThrTyr 1615
4127 AACGGGCGACACATTTTCATCAGCCTATTGAGCCTGCTT... 4173
      ::::: :::::||||| |||||
1615 IsteLgylGlu...PheSerValThrProIleLeuSerAlaArgTyrAsp 1630
4174 ACCGATCCCGCTTCGGGCAAGTCGACACAGCGCTCAATACGCCCTATT 4223
      ::::: :::::|||||
1631 ThrAnsGlnGlySerGlyLysIle.....AsnValAsnGlnTyrAsp 1645
4224 GGCTCAGATTTTCGGCAAAACCCGCAATGGCGGATGGCGGTAACGCCG 4273
      ::::: :::::|||||
1645 eAlaTyrAsnValGlnAsnGlnGlnGlnTyrAsnAlaGlyLeuLysLeu 1662
4274 AAATCAAGGTTTCACGCTGCTCCACGCTCCGCGCAAGGCCG 4323
      ::::: :::::|||||
1662 ystYrHisAsnValLysLeuSerLeuIleGlyLeuThrLysAlaLys 1678
4324 CAACGTGAAGCGACACACAGCGCGGCAATCAATTAAGGCTAC 4365
1679 GlnAlaGlnLysGlnLysThrAlaGlnLeuLysLeuSerPhe 1692
seq_name: p1r2:A41859

```

```

seq_documentation_block:
19A-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
C:Species: Haemophilus influenzae
A:Variety: strain HK715
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
A:Accession: A41859
R:Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A:Reference number: A41859; MUID:92234949
A:Accession: A41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1702 <POU>
A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907
A:Experimental source: strain HK715
A:Note: Sequence extracted from NCBI backbone (NCBI:97282)
C:Superfamily: 19A-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

alignment_scores:

```

Quality: 1566.00      Length: 1872
Ratio: 1.671          Gaps: 49
Percent Similarity: 50.053      Percent Identity: 24.947

```

alignment_block:

```
US-09-303-518D-649 x A41859 ..
```

```
Align seg 1/1 to: A41859 from: 1 to: 1702
```

```

64 CGCTTCGCGCTGCTTACTGACCATATGCTGCTGCTGCAATTTCTTC 113
      ::::: :::::|||||
5 LysPheLysLeuAsnRheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAAGCTGGGCGGACACACTTATTCGGCATCACTACCAATATATATC 163
      ::::: :::::|||||
21 GtYrThrGlnAlaLeuValArgAspAspValAspTyrGlnIlePhe 38
164 GCGACTTGGCGCAAAATTAAGCGAAGTTTGAGTCGGGCGGAAAGATAT 213
      ::::: :::::|||||
38 rGAspRheAlaGlnAsnLysGlnTyrPheSerValGlyAlaThrAsnVal 54
214 GAGGTTTCAACAACAAAGGAGCTGTCGCAATCAATCAACAAAGC 263
      ::::: :::::|||||
55 GlnValArgAspLysAsnAsnHisSerLeuGlnLysAsnValLeuPro 71
264 C...CGGATGATGATTTTCTGCTGCTGCG...CGTAACGGCGTGGCG 307
      ::::: :::::|||||

```

```

71 yLleProMetIleAspPheSerValValAspValAspLysArgIleAla 88
308 CATGGTGGGCGAATCATATATGTTGAGCGTGGCAAT...AACGGCGC 354
      ::::: :::::|||||
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATACACAGCTGATTTTGGT.....GCCGAGGAAGANA 389
      ::::: :::::|||||
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 121
390 TCCGATCAACATCGTTTACT.....TATAAATTG 421
121 AspLysSerHisArgAspValSerSerGlnLysAsnArgTyrPheSer 138
422 TGAACGGAATTAATTAAGCAGGAGCTAAAGC..... 456
      ::::: :::::|||||
138 alGlnLysAsnGlnTyrProThrLysLeuAsnGlnLysAlaValThr 154
457 .....CATCCTTAGCGGCGATTAATCATATGCGCGCTTGA 494
155 GlnAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgLeu 171
495 TAAATTGTCAGATGACAGAACCTGTGA...ATGACCACTTATATG 541
171 pLysPheValThrGlnValAlaProIleGlnLysSerThrAlaSerSer 188
542 ATGGCGGCAATATATGATCAAAATTAATTAATTAATTAATTAATTA 591
      ::::: :::::|||||
188 spAlaGlnThrTyrAsnAspGlnAsnLysTyrProAlaPheAlaArg 204
592 GGGGAGCGAGCAATATGCGGATCTGATGAAGTGAAG..... 630
205 GlySerGlnSerGlnPheIleTyrLysLysGlnAspAsnTyrSerLeu 221
631 .CCCAATACCGCGAAAGTTTCATATCAT.....ATTGCAAGTG 667
221 eLeuAsnAsnHisGlnValGlyLysAsnAsnLeuLysLeuValGlyAsp 238
668 CGATTTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
238 lArgTyrThrTyrGlnLeuAlaGlnLysProTyrLysValAsnHisGln 254
718 GGTGGCACAGTCACCTAGTAGTAAGAAATTAACATAGC...CCATA 764
255 AsnGlnLysLeuIleGlyPheGlyAsnSerLysGlnLysSerAspPro 271
765 TGGTTTTTACCACACAGA.....GCTCATTTTGGCGACA 799
271 scLysIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGly 288
800 GTGGCTCAACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAT 849
288 eArgLysSerProLeuPheValTyrAspArgGlnLysGlyLysTyrLeu 304
850 AATGGGATATTGCAACCGGCAACCCCTATATAGAAAGCAATGGCTT 899
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGln 321
900 CCAGCTGTTGCTAAAGATTGCTTATGATGAATTTTCTGGAGATA 949
321 uTPAsnLysLysProGlnPheAlaLysThrValLeuAspLysAsp 338
950 CCATTCAGTATTTCTAGAACACAGTCGCAAAATGGAATCTGTTTAA 999
338 hrAlaGlnLysLeu.....ThrGlySerAsnThrGlnTyrAsnTrp 352
1000 GACGATATATATGCGACAGCAAAATCAATGCAACATGACCAATTC 1049
353 Pro.....ThrGlyLysThrSerVal..... 359
1050 TCTGCTAATATGATTAATAACAGCAACCGCTCAATGTTTATGTTTCT 1099
360 .IleSerAsnGlnLysSerGlnSerLeuAsnValAspLeuPheAspSer 376

```

1100	TATCCGACGACGACGAAGAACCTGTTATATCATGCTGACGCTGTCACAC	1149
1101	
376	lAspPthAspSerfs.....	381
1150	AGTTATCGACCGACAGCTCATATAATGAGAAATATTCCTTATTTGACGA	1199
382 LysAsnAsnHisGlyLysSerValThrLeu.....	At 392
1200	AGGAAAGCGCATTTGATTACTTACCAGCACATCATATCAAGCTGCGAG	1249
392	gGlySerGlyThrLeuThrLeuAsnAsnAsnHisLeaspGlnGlyValGlyG	409
1250	GATTATATTTCCACAGGACGATTTTTCAGGTC...TCGCTGAAATTAACGA	1296
409	LysLeuPhePheGlnUolLysAspTyrGlnValLysGlyThrSerAspSerThr	425
1297	ACTTTGCAAGGCGCGGGCGGCTTCATATACGTGAAAGACAGTACCGTTACTTG	1346
426	ThTrpLysGlyValGlyValSerValAlaAspLysThrValThrTrp	442
1347	GAAAGTAAACGGCGTGGCAACAGCGCGCTGTCCAAATATGGCCAAAGCA	1396
442	PLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLysGlyT	459
1397	CGCGTACGACGTCAGACGCAAGGCGAAACCAAGGCTCGATCAGCGTGGC	1446
459	hTrpLeuValGlnGlyLysGlyLysAsnLysGlySerLeuLysValGly	475
1447	GACGGTACAGTCACTTTTGGATGACGACGACAGCAGTAAAGCCAAATAACA	1496
476	AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysValLys	492
1497	AGCCTTAGTAAGAAATCGCTGTGTCAGCGGACGAGGGGTACGGTGCACATGA	1546
492	sAlaPheSerGlnValGlyIleValSerGlyArgSerThrValValLeu	509
1547	ATGCCGATTAATCAGTTCAACCCCGACAAACTATATTTGGCTTGGCGGC	1596
509	sAspAspLysGlnValAspProAsnSerIleTyrPheArgLysGly	525
1597	GGAGCTTGATTTAAACGGGCGCTTCGCTTGGTTCACCGCTATTTCAAA	1646
526	GlyTyrLeuAspAlaAsnGlnLysAsnAsnLeuThrPheGlnHisIleArgAs	542
1647	TACCGATGAAGGGCGCATGTCATCAACCAATCAAGACACAGATATCA	1696
542	nIleAspAspGlyValArgLeuValAsnHisAsnThrSerLysThrSerT	559
1697	CCGTTACCATTTACAGGC.....	1713
559	hTrValThrIleThrGlyLysLeuSerLeuIleThrAspProAsnThrIleThr	575
1713	1713
576	ProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaIleArgArg	592
1713	1713
592	gIleLysAspLysGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyrT	609
1713	1713
609	YrAlaLeuArgLysGlyAlaSerThrArgSerLeuLeuProLysAsnSer	625
1714ATAAAGATATTCG	1722
626	GlyGlnSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAspGluAl	642
1728	TACAACGGGCAATACAACAGCTTGATACCAAAAAGAAATTCGGCTACA	1777
642	aLysArgAsnValLeuSerAsnHisIleAsnAsnGluArgMetAsnGlyPheA	659

1778	ACGGTGGTTTGGCGAGAAAGATACGACCAAAAGCAAGGGCGGCTGCAC	1827
659	 snglytyrpneglyglu glu ... glylysasn sngly asn leu asn	674
1828	CTGTGTTACAGACCGCGCGCGAGAGACCGACCCCTGCTGCTTTCGCGGG	1877
675	val thr phe ys s gly lys ser glu glu asn trp phe leu thr gly l	691
1878	AACAATTTTAAAGCGCAACATCGACCAAAACAAGCGCAACTGTTTTTCA	1927
691	 ythr asn leu asn gly asp leu lys val glu lys glu thr leu phe leu s	708
1928	CGCGCACACCAACACCGGACCGCTCAAT.....	1956
708	 erg ly arg pro thr pro his ala trp ile al gly l le ser ser thr	724
1957	CATTAAACGACCAATTGGTCGCAAAAAGAGGCACTTCTCCGCGGGAAT	2006
725	 lys lys asp glu his phe al glu asn asn glu va	736
2007	CGTGTGGACACACATCGATCAACCGGACATTTAAGCGGAAACTTCC	2056
736	 l val val glu asp asp trp ile asn trp asn phe lys ala trp asn lle a	753
2057	AAATTAAGGCGAGCGAGCGCTGTTC...CGCAATGTCGCAAAAGTG	2103
753	 s nval thr asn asn al thr leu tyr ser glu Arg asn val ala sn lle	769
2104	AAAGCGCATTTGGCATTTGACATACGCGCAAGAGATTTTGGCTCGC	2153
770	 thr ser asn lle thr al le ser asp asn al lys val his lle gly tyr	785
2154	ACCGCATAAAGCCACACAAATCTGACACGTTTCGCACTGACGCGTCTGA	2203
786	 ... lys al glu asp trp thr val cys val arg ser asp tyr thr glu tyr	801
2204	CAAAATTGTGCGAAAAAACCATTACCGAGCATTAAGTCATTTGGTTCATGG	2253
801	 al thr cys thr thr asp lys leu ser... asp lys al leu asn ser phe	816
2254	ACTAAGCCGACATCGACGCGCAATGTCGATTCGCGCATCGCTCATTT	2303
817	 asn al al thr asn val ser glu asn val asn leu ser.....	828
2304	AAATCTCACAGGCGTTGCCACACTCAAGCGCAATCTTAGTGCAAATGGCG	2353
828	828
2354	ATACAGCTATACAGTCAGGACGACCAAGCCACCCAAAAGCGCAACTTAGC	2403
828	828
2404	CTTCGTGGCAATGCCAAGCAACTTTAATCAAGCCACATTAAAGCGCAA	2453
829	 ... gly asn al asn phe val leu gly lys al asn leu phe gly th	843
2454	CACATGCGCTGGGGAAGATGCTCATTTATTAAGAGGACGACGCGCTAC	2503
843	 r le ser gly th thr le asn ser glu val al gly leu thr glu.....	856
2504	AAAGCGCAGTCTGACGCTTTCGCGCAAGCGCTAAGCAAAAGCTAAGCAT	2553
856	856
2554	TCCGCACTCAAGGTAATGTCCTCCCTAGCGGATTAAGGACAGTATTCATTT	2603
856	856
2604	TGAAGACGCGCTTTACCGGACAAATCAGCGCGGCAAGGATACGGCAT	2653
856	856
2654	TACACTTAAAGACAGCGAATGACGCTGCCGCTAGACGACGGAATTAAGC	2703


```

3562 ACGTCAACAGCGCTTTTCGCCGTACAGAGCAATTAGACCGGTA...TT 3608
      |||||
1433 AAldeuasnValGlyLysAlaValSerGlnHisLeSerGlnLeuLume 1449
      |||||
3609 TGGCGAGACCGCGCGACCGCTTTGGACAGCGGCATCCGGGACACA 3658
      |||||
1449 tAsnAsnGlnGlyGlnThrValSerValSerAsnThrSerMetAsnG 1466
      |||||
3659 AACACTTCGTCGCAAGATTTCCGCGCTACCGCCCAACAAACGACCTG 3708
      |||||
1466 lUsnThrSerSerSerGlnThrArgPheSerSerLysSerThrGln 1482
      |||||
3709 CGCGAAATCGGTATGCAGAAAACCTCGGACGGCG...CGCGTCGGCAT 3755
      |||||
1483 ThrGlnLeuGlyThrPheArgGlnThrIleSerAsnValGlnLeuGly 1499
      |||||
3756 CCTGTTTCCGACACCGGACCGAAACACCTTCGACGACGCGCATGCGCA 3805
      |||||
1499 yValPheThrThrValAlaPheSerAsnValPheAspLysAlaSerSerL 1516
      |||||
3806 ACTCGGACGCGCTTCCCGACGCGCGCTTTCGGGCAATACGCGCATCGAC 3855
      |||||
1516 ySAsnThr...LeuAlaGlnValAsnPheThrSerLysThrValAlaAsp 1531
      |||||
3856 ...AGGTTTACATCGCGCATCGACGCGCGCGGCTTTAGCAGCGCGCAG 3902
      |||||
1532 AsnHisThrPheGlyLeuGlyIleAspLeuGlyThrValLysPheGlnSerAs 1548
      |||||
3903 CCTTTCAGACGCGCATCGAGGCAAAATCCCGCGCGCGCTGCGCATACG 3952
      |||||
1548 nLeuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheG 1565
      |||||
3953 GCATTTCAGGCGCGATACCGCGCGCTTTCGGGATTCGGATCGACGACG 4002
      |||||
1565 lLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrPro 1581
      |||||
4003 CACATCGCGCGACGCGCTATTTCGCCAAAAACGCGATTACCGCTACGA 4052
      |||||
1582 lLevalGlyValAlaArgThrSerThrLeuSerAsnHisAlaPheAla 1598
      |||||
4053 AAACGTCATATCGCCACCGCGCGCTTTCGATTCACCGCGTACCGCGCG 4102
      |||||
1598 aLysAspArgIleLysValAsnProIleSerValLysThrAlaPheAla 1615
      |||||
4103 GCATTAAGGAGATATTCATTCACAAACGCGGCAACACATTCATTCACG 4152
      |||||
1615 lValAspLeuSerThrThrThrHisLeuGlyGlu...PheSerValThr 1630
      |||||
4153 CCTATTTCAGCGCTGCTCTAT...ACGATGCGCGCTTCGGCGAAAGTCCG 4199
      |||||
1631 ProIleLeuSerAlaArgThrAspThrAsnGlnGlySerGlyLysIle 1646
      |||||
4200 AACACGCGCATACCGCGCTTTCGCGCATTCGCGCAAAACCGCGCA 4249
      |||||
1647 ...AsnValAsnGlnThrAspPheAlaThrAsnValGlnHisGlnGlnG 1662
      |||||
4250 GTGCGGAATGCGCGTAAACGCGCAATCAAAAGTTTCACGCTGTCCTC 4299
      |||||
1662 lThrValAsnAlaGlyLeuLysLysLysLysLysLysLysLysLysL 1678
      |||||
4300 CACGCTGCGCGCGCAAGCGCGCGCATCGGAAAGCGCAACACGCGCGG 4349
      |||||
1679 lGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG 1695
      |||||
4350 CATCAATTAAGGCTAC 4365
      |||||
1695 uLeuLysLeuSerPhe 1700
      |||||
seq_name: p1r2:A81018
seq_documentation_block:
serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogroup

```

```

C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81018
R:Retellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Halt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000, MID:20175755
A:Accession: A81018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <TE1>
A:Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAF42325.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1998

```

```

alignment_scores:
Quality: 1547.00 Length: 1669
Ratio: 1.809 Gaps: 50
Percent Similarity: 51.228 Percent Identity: 27.741

```

```

alignment_block:
US-09-303-518D-649 x A81018 ..

```

```

Align seg 1/1 to: A81018 from: 1 to: 1431

```

```

55 GCGCGCATCGCTTCGCCGTCTACTTACGCAATGCGCTGCTGCG 104
      |||
17 GlyLeuPheAlaValSerProAlaThrSerSerIleValAlaAsn.... 31
      |||
105 CATTTTCCCAAGCCTGGCGGACACACTTATTTCGCGCATCACTACC 154
      |||
32 .....AspValAspThrG 36
      |||
155 AATACTATGCGCAGCTTTCGCCGAAATTAAGCGACGTTTCAGTCGGGCG 204
      |||
36 lThrPheArgAspPheAlaGlnAsnLysGlyAlaPheThrValGlyAla 52
      |||
205 AAGCATATTGAGGTTTACAAACAAAGGAGGAGTTGTCGCGCAATTCAT 254
      |||
53 SerAsnHisSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69
      |||
255 GACAAAAGCGCGATGATGATTTTCTGCTGCTGCGCT...AACGCGG 301
      |||
69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86
      |||
302 TGGCGGATGCGGCGGATCATATATGTCGCTGCGCATACGAC 351
      |||
86 lValThrLeuValHisProGlnThrValAsnSerValLysHisAsnVal 102
      |||
352 GCGTATTAACACGTTGATTTGGTCGCGAAGGAAATCCGATCAACA 401
      |||
103 GlyThrGlySerIleGlnPheGlyAsnAspThrGlnHisProGlnGlnG 119
      |||
402 TCGTTTACTTATTAATTTGTGAACGAGATATTTAAGCAGGGACTA 451
      |||
119 nAlaThrThrThrArgLeuValSerArgAsn..... 129
      |||
452 AAGCGCATGCGTTATGCGGATTCATATGCGCGCTTTCGATTAATTT 501
      |||
130 ...ProHisPro...AspThrAspThrLysLeuProAlaGlnLysLeu 144
      |||
502 GTCAACAGATCGAAGCTGTGAATGACAGCT..... 534
      |||
145 ValThrGlnLysSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161
      |||
535 ...TATATGATGCGGCGGAAATATATGATCAAAATATATACCTGACC 580
      |||
161 nGlyGlnProLysAlaAsnAlaThrLysAspThrAspArgPheProThrP 178
      |||

```

581 GTTTCGATTGGGCGAGCGAGCAATATGCGATCTGATGAGATGAG 630
178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193
631 CCCAATAACCGGAAAGCTTCATATCATATTTGCAAGTCCGATATCTTGGCT 680
194ThArgThrArgThrAlaProAlaTyrGlnTyrLe 205
681 CGTTGGTGCATATCCTTTGCACAAATGGATGCGTGGTGGCGACAGTCA 730
205 uThcLysGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222
731 ACTTAGGTAGTGAATAATTAACATAGCCCATATGGTTTATTAACCA 780
222 euValGlyGly...SerLeuThrAspLysPro.....LeuAsnThr 234
781 GGAGGCGCATTTGGCGAGTGGCTCCCAATGTTATCTATGATGCCCA 830
235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysH 251
831 AAAGCAAAAGTGTATTAATTAATGCGGTATGCAACGCGCAACCCCTATA 880
251 sGlnAsnArgTyrValLeuAlaGlyValLeuSerThr.....TyrA 265
881 TAGGAAAGCAATGGCTCCAG.....CTGGTGTAAAGATTGGTTC 924
265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281
925 TATGATGAATCTTGGCTGAGATACCATTCGATTTCTAGCAACCCAG 974
282 lLeArgSerThrIleArgGln.....TyrGlnThrArg 292
975 T.....CAAAATGGAAATACCTTTTAACGACGATTAATA 1009
292 gLeuAspValGlyLeuThrThrAsnGlnLeuIleTyrArgAspAsnGlyA 309
1010 ATGGCAGAGAAATCAATGCCAATGACACACATCTCTGCTTAAT 1059
309 snGlyAsnSerThrLeuGlnGlyLeuAsnGlnArgIleThrLeuPro... 324
1060 AGATTAAAAACGAAACCGTCAATGTGTTATGTTCTTTATCCGAGAC 1109
325IleAlaAsnProSerLeuAla..... 331
1110 AGCAAGAGAACTGTTATCATGCTGCAGGTGGTGTCAACAGTTATTCAC 1159
332P 332
1160 CCAGACTGAATATGAGAAATATTTCTTATGACGAGGAA... 1206
332 roGln...AsnAspSerArgHisMetProSerGlnAspAlaGlyLysThr 347
1207GGCGAATTGATTAATCTTACACAGAA 1229
348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364
1230 CATCAATCAAGTGTGCTGAGATATTAATTCAGAGGAGATTTACGGTCT 1279
364 nLleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrValV 381
1280 CGCCTGAAATTAACGAAACTTGGCAGGCGGCGCTTCATATCAGTAA 1329
381 aGlyLysAsnHis...ThrTyrGlnGlyAlaGlyValIleValAlaAsp 396
1330 GACATACCGCTTACTTGAAGTAAAGCGCTGCGCAACGACGCGCTGTC 1379
397 GlyLysArgValPheTyrPheGlnValSerAsnProLysGlyAspArgLeu 413
1380 CAAATTCGCGAAAGCAGCAGTTCAGGCAACCAAGGAAAGCAACCAAG 1429
413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGlnG 430
1430 GTCGATACGCGTGGCGAGCGGTACAGTCAATTTTGATACGACGAGAC 1479
430 lYAspIleSerIleGlyGlnGlyThrValValLeuAlaGlnLysAlaAla 446
1480 GATAAAGCAAAAACAGCCTTAGTGAATAGCGCTGGTGCAGCGCAG 1529
447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyArg 463
1530 GGGTACGGTGCACCTGAATGCCATTAATCAGTTCAACCCGCAAACTCT 1579
463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGlnAsnLeuT 480
1580 ATTTCGGCTTTCGCGCGGACGTTTGGATTAAACGGGCAATTCGCTTTCG 1629
480 yPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlnLysAsnLeuAla 496
1630 TTCACGCTATTCAAATTAACGATGAAGGGCGCATATGTCACACAA 1679
497 PheThrHisIleArgHisAlaAspGlyGlyAlaGlnIleValAlaAsnHis 513
1680 TCAAGACAAAGATCCACCGTTCACATTAACGAGCAATTAAGATATGCTA 1729
513 nProAspGlnAlaAlaThrLeuThrLeuThrArgLysAsnProValLeuSer 530
1730 CAACC..... 1734
530 roGlnHisValGluTyrValGlnTyrPheGlyAsnArgProGlnGlyAsnAla 546
1734 1734
547 AlaValTyrGluTyrIleAsnProHisArgAsnArgAlaGlyThrAspTyrPhe 563
1735GCGCAAT..... 1740
563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580
1741AACAACGCTTGATAGCAAA 1761
580 yAsnSerThrSerTyrPheGlnPheIleGlyAsnAsnArgGlnIleAlaAla 596
1762 AAAGAAATGGCC.....TACAA 1778
597 GluGlnValAlaGlnAlaGlnAsnAlaArgProAspLeuIleThrPheG 613
1779 CGGTGGTGTGGCGGAAAGATACGACC..... 1806
613 yGlyTyrLeuGlyGlyAsnAlaGlnThrGlyLysAlaAlaProSerTyr 630
1806 1806
630 erLysThrAsnGlnAlaAlaIleGlyLysThrArgHisIleAlaAsnAla 646
1807AAACGAAAGCGGCGCTCAACCTTGT 1832
647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuH 663
1833 TTACAGCCGCCCGCAGAGACCGACCGACCTGCTTTCGCGGAGCAAA 1882
663 sTyrArgProLysArgThrAspSerThrLeuLeuAsnGlyGlyMetA 680
1883 ATTTAAAGCGCAACATCAGCAAAACGCGCAAACTGTTTTCAGCGG 1932
680 snLeuAsnGlyGlyValLeuIleGlnGlyLysAsnMetIleValSerGly 696
1933 AGACCAACGCGCACGCTTACATCAATTAACGACCATGTCGCGCAAA 1982
697 ArgProValProHisAlaTyr.....AspHisGlnAlaLysArg 709
1983 AGAGGGCATTCCTCGCGGGAATCGTGTGCGCAACGATCGATCAAC 2032
709 gLysPro.....ValLeuGlnAsnGluTyrThrAspG 720
2032 GCACATTTAAAGCGGAAACTTCGAAATTAAGGC...GGACAGGCGGTG 2079

```

      :::::::::: ::::: ::::: ::::: :::::
720 lysesrPheLysAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736
      :::::::::: ::::: ::::: ::::: :::::
2080 GTTCCCGGCAAGTGTGCCAAAGTGAAGGCGAT.....TGCCATT 2120
      :::::::::: ::::: ::::: ::::: :::::
737 AlagIyArgAsnThrAlaHisLeuAspGlyAspIleThrAlaTyrAspLe 753
      :::::::::: ::::: ::::: ::::: :::::
2121 GAGCAAT...CACGCCAAGCATTTTGTGTGTCGACCGCATCAAAAGCC 2167
      :::::::::: ::::: ::::: ::::: :::::
753 uSerGlyIleAspLeuGlyPheThrGlnGlyLysThrProGln..... 767
      :::::::::: ::::: ::::: ::::: :::::
2168 ACACAAATGTGACAGCTTGCGAGCTGAGCGGCTGTGACAAATGTGTGCA 2217
      :::::::::: ::::: ::::: ::::: :::::
768 .....CysTyrArgSerTyrHisSerGlySerThrHisCysThrPro 781
      :::::::::: ::::: ::::: ::::: :::::
2218 AAACCAATTCACGACGATAAAGTGAATGTTCATTGTACTAGACCGCAT 2267
      :::::::::: ::::: ::::: ::::: :::::
782 AsnAlaValLeuLysAlaGlnAsnTyrArgAlaLeuProAlaThrGlnVa 798
      :::::::::: ::::: ::::: ::::: :::::
2268 CAGCGGCAAGTGCATCTTGCCGATCGCTCATTTAAATCTCACAGGCG 2317
      :::::::::: ::::: ::::: ::::: :::::
798 lArgGlyAspIleThrLeuAsnAspArgSerGluLeuArgLeu...GlyL 814
      :::::::::: ::::: ::::: ::::: :::::
2318 TTGGCAACCTCAAGCGCAATCTTAGTCAATGGCATACAGCTTATACA 2367
      :::::::::: ::::: ::::: ::::: :::::
814 ySalAlaHisLeuTyrGlySerIleArgAla..... 823
      :::::::::: ::::: ::::: ::::: :::::
2368 GTCAAGCCAAAGCCACCCAAAGCGCAACCTTAGCCTCGTGGCAATGC 2417
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2418 CCAAGCAACATTAAATCAAGCCACATTAAAGGCAACATCGGCTCG 2467
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2468 GCAATGCTCATTTAATCTAAGCGACCGGCTACAAAGGCGAGTGTG 2517
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2518 ACGTTTCCGGCAACGCTAAGCGAAAGTAAAGCCATCCGACATCAAGC 2567
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2568 TAATGTCTCCCTAGCCGATAAAGCAGATTTCCATTTTGAAAGCAGCCGT 2617
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2618 TTACGGGACAATCAAGCGGCGCAAGATACGGCATTTACACTTAAA... 2664
      :::::::::: ::::: ::::: ::::: :::::
824 .....GlyLysAspThrAlaValAlaArgMetGluAla 833
      :::::::::: ::::: ::::: ::::: :::::
2665 GACAGGGAATGAGCGCTGCCCTCAGAGCAAGGGAATTAGCAATTTAAACT 2714
      :::::::::: ::::: ::::: ::::: :::::
834 AspSerAsnTyrThrLeuSerGlnSerSerHisThrGlyAlaLeuThrLe 850
      :::::::::: ::::: ::::: ::::: :::::
2715 TGCAACAGCGCAACATTACACTCAATTTCCGCTATCGCCAGCATGGCGAG 2764
      :::::::::: ::::: ::::: ::::: :::::
850 uAspGlyAlaGlnIleThrLeuAsnProAspPheAlaAsnAsn..... 864
      :::::::::: ::::: ::::: ::::: :::::
2765 GGGCGCAAAACGGGAGTGGCAGACAGATGGCGCGCGCGCTTCGGCGCT 2814
      :::::::::: ::::: ::::: ::::: :::::
864 ..... 864
      :::::::::: ::::: ::::: ::::: :::::
2815 TCGCGCGCTTCCTATTATCCGTTACACGCCCACTGCGTGAATCCG 2864
      :::::::::: ::::: ::::: ::::: :::::
865 .....ThrHisAsnAsnArg 869
      :::::::::: ::::: ::::: ::::: :::::
2865 TTTCACAGCGCTGACGTAAGCGCAATTTGAAGCGTACAGGACATTTCC 2914
      :::::::::: ::::: ::::: ::::: :::::
869 gPheAsnThrLeuThrValAsnGlyThrLeuAspGlyPheGlyThrPheA 886
      :::::::::: ::::: ::::: ::::: :::::
2915 GCTTTATGTGGAACTCTTC...GGCTACCGCAGCGCAAAATTGAAGCG 2961
      :::::::::: ::::: ::::: ::::: :::::

```

```

886 rGPheLeuThrGlyIleValAlaArgLysGlnAsnAlaProProLeuLysLeu 902
      :::::::::: ::::: ::::: ::::: :::::
2962 GCGGAAGTTCGGAAGCAGCTTACACCTTGCGGCGTCAACATTAACGCGCA 3011
      :::::::::: ::::: ::::: ::::: :::::
903 GluGlyAspSerArgGlyAlaPheGlnIleHisValLysAsnThrGlnGly 919
      :::::::::: ::::: ::::: ::::: :::::
3012 CGAACCTGCAAGCGCTCGAACAATTCAGCGTACTGAGAGAAAGCAACA 3061
      :::::::::: ::::: ::::: ::::: :::::
919 ngIuProGlnThrThrGlnSerLeuAlaLeuVal.....SerLeuAsn 934
      :::::::::: ::::: ::::: ::::: :::::
3062 AACCGCTGCCGAACCTTAATTTCAACCCGCAAAAGCAACGCTCAT 3111
      :::::::::: ::::: ::::: ::::: :::::
934 roLysHisSerHisGlnAlaAlaArgPheThrLeuGlnAsnGlyTyrAlaAsp 950
      :::::::::: ::::: ::::: ::::: :::::
3112 GCGGCGCGGCGGCTTACCAACTCATCCGCAAGAGCGGCGACTTCGCT 3161
      :::::::::: ::::: ::::: ::::: :::::
951 LeuGlyAlaTyrArgTyrIleLeuArgLysAsnAsnAsnGlyTyrSerLe 967
      :::::::::: ::::: ::::: ::::: :::::
3162 GCATTAATCGGTCAAAGAACAGAGCTTCCGCAAACTCGGCAAGGCG 3211
      :::::::::: ::::: ::::: ::::: :::::
967 uTyrAsnProLeuLysGluAlaGluLeuGlnIleGluAlaThrArgAlaG 984
      :::::::::: ::::: ::::: ::::: :::::
3212 AA.....GCC 3216
      :::::::::: ::::: ::::: ::::: :::::
984 lHisGluAlaArgAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIle 1000
      :::::::::: ::::: ::::: ::::: :::::
3217 AAAAACAAGCGCGGAAAAAGACAAC..... 3240
      :::::::::: ::::: ::::: ::::: :::::
1001 SerArgGlnValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAl 1017
      :::::::::: ::::: ::::: ::::: :::::
3240 ..... 3240
      :::::::::: ::::: ::::: ::::: :::::
1017 aTrpGlnAsnSerGlnThrGluLeuAlaArgIleAspSerGlnValGlnT 1034
      :::::::::: ::::: ::::: ::::: :::::
3240 ..... 3240
      :::::::::: ::::: ::::: ::::: :::::
1034 yrLeuSerAlaGlnLeuLysGlnThrAspProLeuThrGlyIleLeuThr 1050
      :::::::::: ::::: ::::: ::::: :::::
3241 ..GCGCAAGCTTTGACGCGCTG..... 3261
      :::::::::: ::::: ::::: ::::: :::::
1051 ArgAlaGlnAsnLeuLysAlaAlaGlnGlyTyrSerAlaAspIleCysAr 1067
      :::::::::: ::::: ::::: ::::: :::::
3262 ....ATTGCGGCGCGGCGCGATGCC..... 3282
      :::::::::: ::::: ::::: ::::: :::::
1067 gGlnValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGluThrC 1084
      :::::::::: ::::: ::::: ::::: :::::
3283 .....GTGAAAAAGACAGAAAGCGTTGCCGA 3309
      :::::::::: ::::: ::::: ::::: :::::
1084 lLeuAspThrTyrIleGluArgValGluMetAlaGlnSerGluLeuAsp 1100
      :::::::::: ::::: ::::: ::::: :::::
3310 CCGGCGCGCGACGCGCGGCGGAAATGCGCATTTAGCAGCGCGAGCA 3359
      :::::::::: ::::: ::::: ::::: :::::
1101 LysAlaAlaArgGlnGlyLysPhe.....AlaGlnAlaValGln 1112
      :::::::::: ::::: ::::: ::::: :::::
3360 AGAGAAAAAAGCGGTGACGGCGGATTAAGACACCGCTTGCGCAAGACAG 3408
      :::::::::: ::::: ::::: ::::: :::::
1112 uThrAlaArgHisAlaTyrLeuAsnAlaLeuAsnArgLeuSerArgGlnI 1129
      :::::::::: ::::: ::::: ::::: :::::
3409 ..CGCGAAGCGGAACCGCGCGGTACACCGCTTCCCGCGCGCGC 3456
      :::::::::: ::::: ::::: ::::: :::::
1129 lHisSerLeuLysThrGlyValAlaGlyIleArgMetProAsnLeu... 1144
      :::::::::: ::::: ::::: ::::: :::::
3457 CCGCGCGCGCGGAGTTTGGCGCACTGCACCCCAACCGCAGCCCAACC 3506
      :::::::::: ::::: ::::: ::::: :::::
1144 ..... 1144
      :::::::::: ::::: ::::: ::::: :::::
3507 GCAGCGGACCTGATCAGCGCTTATGCCAATAGCGGTTTGAGTAATTT 3556
      :::::::::: ::::: ::::: ::::: :::::
1145 ....AlaGluLeuIleSerArgSerAlaAsnThrAlaValSerGlnGln 1160
      :::::::::: ::::: ::::: ::::: :::::
3557 CCGCCACGCTCAACAGCGTTTCCGCGTACAGAGCAATTTAGCGCGCTA 3606
      :::::::::: ::::: ::::: ::::: :::::
1160 lAlaTyrAsnThrThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHis 1176

```

```

3607 TTTCGCAAGACCCGCGCAAGCGGTTTGACAGCGGCAATCGGAGAC 3656
      :::::::::: :::::::::: ::::::::::
1177 LeuThrSprProGlnGlnAsnIleThrPheGluThrGlyThrGln 1193
      :::::::::: ::::::::::
3657 CAACACTACCGTTCGCAAGATTTCCGCGCTACCGCAACAAACGAC 3706
      :::::::::: ::::::::::
1193 nThrSprGlnHisSerGlyThrHisArgProGlyGlnGlnThrAsn 1210
      :::::::::: ::::::::::
3707 TCGCCCAATCGGTATGCAAGAAAACCTCGGACGCGG...CGCGTCGC 3753
      :::::::::: ::::::::::
1210 yralahistilegilyleGlnThrGlyleThrSprArgLeuSerValGly 1226
      :::::::::: ::::::::::
3754 ATCCGTGTTTCGCAACCGGACCGCAACACTTCGACGAGCGATCGG 3803
      :::::::::: ::::::::::
1227 ThrIleLeuThrSprGlnArgThrAsnAsnArgPheAspGlyValSe 1243
      :::::::::: ::::::::::
3804 C.....ACTCGGACGCGCTTCGCGCGCGCGCTTTGGGCAATACG 3847
      :::::::::: ::::::::::
1243 rAlaArgAsnArgSerAsnGlyAlaHisIleThrValIlysglyluAsn 1260
      :::::::::: ::::::::::
3848 GCATCGACAGAGTTCTCATCGCATCGCGCGCGCGGCTTTTGACGC 3897
      :::::::::: ::::::::::
1260 lyla.....LeupheAlaAlaAlaAspLeuGlyTyrSerAsn 1272
      :::::::::: ::::::::::
3898 GCG.....AGCTTCAGACGCGCATCGGAGCAAAATCCGCGCGCGCT 3941
      :::::::::: ::::::::::
1273 SerArgThrArgPheThrAspTyrAspGlyAlaAlaValArgGlyHisAl 1289
      :::::::::: ::::::::::
3942 GTCGATTCAGCGATTCAGGCGAGATTCGCGCGCGCTTCGCGGATTCG 3991
      :::::::::: ::::::::::
1289 aThrSprAlaGlyIleAsnThrGlyIleIleuSileasp...ThrGlyIleA 1305
      :::::::::: ::::::::::
3992 GCATCGACCGCAATCGGC.....GCAACGCGC 4020
      :::::::::: ::::::::::
1305 sAlaArgProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArg 1321
      :::::::::: ::::::::::
4021 TATTTGTCGCAAAAGCGGATTCACCGCTACGAAACGTCATATTCGCCAC 4070
      :::::::::: ::::::::::
1322 TyrValLeuAspGlyAlaGlu.....IleAsnSe 1331
      :::::::::: ::::::::::
4071 CCCGCGCTTCGATTCACCGCTACCGCGCGGATTAAGCAGCATATAT 4120
      :::::::::: ::::::::::
1331 rProAlaGlnIleGlnThrThrPheHisAlaGlyIleArgLeuAspTyr 1348
      :::::::::: ::::::::::
4121 CATTCAAACCGCGCGCAACATTCATTCACCGCTTATTTAGCGCTTCC 4170
      :::::::::: ::::::::::
1348 hValAlaGluLeuGlyGln...AlaIlyLeuThrProAlaPheSerSerAsp 1363
      :::::::::: ::::::::::
4171 TAT.....ACCGATGCGCTTCGCGCAAGTCCGACAGCGGCTCATAC 4214
      :::::::::: ::::::::::
1364 TyrTyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAs 1379
      :::::::::: ::::::::::
4215 CGCGCTATTCGCTCAGATTTCCGCAAAACCGGATGCGGAAATGGGCG 4264
      :::::::::: ::::::::::
1379 rArgThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHisThrLeuGln 1396
      :::::::::: ::::::::::
4265 TAAACGCGCAATCAAGGTTTCAGGTCGCTCCGCGCGCTGCGCGCGC 4314
      :::::::::: ::::::::::
1396 leAspAlaGlyTyrIlysglyTyrPheAsnAlaIlysglnAlaIlyGly 1412
      :::::::::: ::::::::::
4315 AAAGCGCGCACTGGAAGCGCAACACACGCGCGGATCAATTAAGGCTA 4364
      :::::::::: ::::::::::
1413 LysAspSerAsnThrAlaArgHisIlysglnAlaGlyIleIlysglyIle 1429
      :::::::::: ::::::::::
4365 CGCGTGG 4371
      ::::::::::
1429 rAsnTTP 1431
      ::::::::::
seq.name: p1r2:A81937
seq.documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - Neisseria meningitidis

```

```

N:Alternate names: Iga proteinase; Iga1 proteinase (EC 3.4.21.7) [misnomer]; immunogl
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain HF117; strain HF159; strain SM1027
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
R.; Holtroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; MUID:20222556
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:97379424; PIDN:CA84182.1; PID:9737
A:Experimental source: serogroup A, strain 22491
R:Lombolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82470; NID:9732854; PIDN:CA457853.1; PID:9732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82471; NID:9732856; PIDN:CA457854.1; PID:9732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82472; NID:9732852; PIDN:CA457855.1; PID:9732853
A:Experimental source: strain SM1027
C:Genetics:
A:Gene: iga; NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
      Quality: 1519.50      Length: 1836
      Ratio: 1.563
Percent Similarity: 52.996      Percent Identity: 27.397

alignment_block:
US-09-303-518D-649 x A81937 ..

Align seg 1/1 to: A81937 from: 1 to: 1773

49 AAAACGCGCGCATCCGCTTCGCGCTTACTTACCATATGCGCTGTC 98
      ||||| :::::::::: ::::::::::
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
      :::::::::: ::::::::::
99 GTTCGCGATTCCTCCCGACGCTGGCGGACACACTATTTGGGATCA 148
      :::::::::: ::::::::::
18 aTyrAlaLeuThrProTyrSerGlyAlaAlaLeuValaArgAspAspVala 35
      :::::::::: ::::::::::
149 ACTGCAATATCTATGCGACTTTCGCAAAATTAAGCAAGTTGCGAGTC 198
      :::::::::: ::::::::::
35 sPtyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
      :::::::::: ::::::::::
199 GGGCGCAAGATATGAGCTTTACACAAAAGGCGAGTTGGTGGCA 248
      :::::::::: ::::::::::
52 GlYAlaThrAspLeuSerValIlysaAsnLysGlnGlyGlnAsnIleGlyAs 68
      :::::::::: ::::::::::
249 ATCAATGACAAAGCCCGATGATGATTTCTGAGTGGTGGCGTAAC. 297
      :::::::::: ::::::::::
68 nAlaLeuSerAsnValProMetIleAspPheSerValaAlaAspValaAsnA 85

```

```
298 .. GCGCTGGCGCATGTGTGGCGATCAATATATTGTGACGCTGCACAT 345
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
85 rgrgrhrleuthralleasprroglintyrallavalserallyshis 101
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
346 AACGGCGGCTATACAACTGATTTGGTGGGAAGAAAGAACCCCA 395
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
102 vallysglyasprleuileserlytyrlyshisashisglyhisleu 118
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
396 T.....CAACATCGTTTACTATATAATTTGAACGGAATATATATA 439
      | :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
118 pvalserashaspriusnslutyrargservalalaglnasaptylg 135
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
440 AACGAGGACTAAAGCCATCTTATGCGGC.....GAT 474
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
135 luproasnlyasasntrhis...hisglyasnslnglyarlgueugluasp 150
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
475 TATCATATGCGCGGTTTGCATATAATTTGCACAGATGCAGAACCTGTGA 524
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
151 tyrasmetalargleuasnlyspnevalthrleuvalalaproileal 167
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
525 AATGACCACTTATATGATGGG...CGGAATATATCATCAAAATATAT 571
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
167 aprothseralaglygllygllyalgluthtyrylysasplyasnarpr 184
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
572 ACCCTGACCGTGTTCATTTGGGCGAGCGCAATATTTGGCATCTGAT 621
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
184 heserleuphevalarvalgllyalgllythrlnphe..... 196
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
622 GAACATGAGCCCAATACCGCAAGTTTCATATATTCATTCAGATGGCGTA 671
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
197 .....glutyrasnserarglytyrasmetthrleuaserarglyaty 211
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
672 TTCTTGGCTCGTGGTGCATATACCTTT.....GCACAAA 706
      | :: :: :: :: :: :: :: :: :: :: :: :: :: ::
211 rargtyrallealaglthrprotyrglnasprvalasnavalthrsera 228
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
707 ATGATCAGTGTGCGCAGCTCACTAGTAGTGAATAATTAACAT 756
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
228 snleuasnglnglyleuileglypheglyasprasnserlyshis 244
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
757 AGCCCA.....TATGTTTTTTTACCACAGG 782
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
245 Serprogluylsleuylsgluvalleuaserlglnasnalaleuthrasnty 261
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
832 AGGCTCATTTGGCGACAGTGGCTACCAATGTTATCATGATGCCAAA 832
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
261 ralaivalleuylasprserglyserproleuphealalarasplysglng 278
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
833 AGCAAAAGTGTATATATGAGGTATTCACAAAGGCGCACCCCTATATA 882
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
278 lulyarqtrvalrphelengllyalarqtrpralagltyr... 293
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
883 GGAATAAGCAATGGCTCCAGCTGGTTCGTAACATTTG...TTCTATGA 929
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
294 ...glnlyasasnsertrpqln.....glutrpasnlietyrly 305
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
930 TGAATCTTTGCTGGAGATACCAATTCAGTATTCACAGCACCGTCAAA 979
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
305 slusleuphealasp.....glulleysglna 315
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
980 ATGGGAATACTCTTTTAAAGCATATATATGCGACAGAAATTCAT 1029
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
315 rgr.....Aspasnalagltythrileysglytyr 324
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1030 GCCAAACATGAACACATCTCTGCTAATAGATTAAAAACAGAACGCT 1079
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
325 glygluhsishstprlysthr.....Thrlythr... 334
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1080 TCAATTTGTTATGTTCTTATCCGACAGACAGAACCTGTTATC 1129
      ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
335 .....Asnerhisileglyserthrhalavalrgleuvalagllya 348
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1130 ATGCTGACAGGTGTGCAACAGTTATGACCCAGACTGAATATAGAGAA 1179
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
```

638 snGlyLysGlyHisnGlyAlaLeuAsnLeuAsnPhesnGlyLysSer 654
1849 GAAGACGGCACCCGTGCTTCCGGGGAGCAAAATTAAAGCGGCAACAT 1898
655 AlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyLys 671
1899 CACGCAAAACAGCGCAACTGTTTTCAGCGCGACACACACCGGACG 1948
671 eSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHis 688
1949 CTTACAAATCATTTAAAGACCATTTGGTCCAAAGAGGCG.....ATT 1992
688 laArgSerPheValAsnLysSerSerAlaGlnLysAspAlaHisPheSer 704
1993 CTTCCGGGGAAATCGTTGGGACAAAGCATGTGATCAACCGCATTTAA 2042
705 LysAsnAsnGlyValAlaPheGlnAspAspTrpIleAsnArgThrPhe 721
2043 AGCGGAAACCTTCCAAATTAAAGCGGACAGCGGCTTCC...CGCA 2089
721 sAlaThrGlnIleThrValAsnGlnSerAlaSerPheSerSerGlyArg 738
2090 ATGTTCGCAAGTGAAGGCGATTGGCATTTGACATCACCGCCACGA 2139
738 snValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaVal 754
2140 GTTTTGGTGTGGACCGGCATCAAAAGCCACACATCTGTACACGTTTCCGA 2189
755 AsnLeuGlyTyr.....LysAsnGlyAspGlyValAlaArgSerAs 769
2190 CTGGACGGGTGTGACAAATTGTGTGAAAAAACCATTCACGACATAAG 2239
769 pTyThrArgLysValThrCysAsnThrGlyAsnLeuSer...AspLys 785
2240 TGATTGCTTCATTGACTAAGACCGACATCAGCGGATGTGATCTTGCC 2289
785 laLeuAsnSerPheGlyAlaThrLysIleAsnGlyAsnValAsnLeuAsn 801
2290 GATCAGCGCATTTAAATCTCAGAGGCTTCCACACTCAAGCGCAATCT 2339
802 GlnAsnAlaAlaLeuValLeu...GlyLysAlaAlaLeuTrpGlyGln 817
2340 TAGTCAAAATGGCGATACAGTTATACGTGACGACACAGCG..... 2382
817 eGlnGlyGlnLysAsnSerArgValSerLeuAsnGlnHisSerLysTrp 834
2383ACCGAAAGCGCAACCTTACGCTC..... 2406
834 isLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHis 850
2407GTGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
851 IleHisLeuAsnAsnAlaSerAspAlaGlnSerAlaAsnGlnLysHis 867
2442 ATTAAACGGCAACACATCGGCTTGGGCAATGCTTCATTTAACTAAGCG 2491
867 fLeuLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisLysTrpLeu 883
2492 ACCAGCGCGTACAAAC...GGCAGTCTGACGCT..... 2523
883 hrHisLeuAlaGlnAsnLeuGlyAspLysValLeuValLysGlnSerAla 899
2524 TCGGCAACGCTAAGCAACGTAAGCCATTCCGACATCAACGGTAAT... 2571
900 SerGlyHisLysGlnLeuHisValGlnAspLysThrGlyLysProAsnGln 916
2571 2571
916 nGlyGlyLeuAsnLeuPheAspAlaSerValAlaArgAspArgSerHisL 933
2572GTCTCCCTAAGCGATTAAGCGATATTCATTTTGAAGACAGCGCG 2616
933 euSerValSerLeuAlaAsnHisVal...AspLeuGlyAlaLeuArg 948
2617 TTTCGCGACAAATCACCGCGCGCAAGATACGGCTTACACTTAAGA 2666
949 TyrThrIleLysThrGlnAsnGly...IleThrArgLeuTyrAsnProTy 964
2667 CACGGAA.....TGACGCTGCCGCTGACGACGCG 2695
964 rAlaGlnAsnArgArgValLysProAlaProSerProAlaThrAsnT 981
2696 AATTAGCAATTTAACTTACACCTTGACACGCCACCATTA..... 2733
981 hrAlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 997
2734CTCAATTCCGCTA 2747
998 IleValValAlaProProSerProGlnAlaAsnGlnAlaGlnValAla 1014
2748 TCGCCACAGATGCGCGAGGGCGCAAAAC..... 2775
1014 sArgGlnGlnAlaLysAlaGlnGlnValLysArgGlnGlnAlaGlnAlaG 1031
2775 2775
1031 luArgLysSerAlaGlnLeuAlaLysGlnLysAlaGlnAlaGlnLysGln 1047
2776GGCAGTGGCAG 2786
1048 AlaArgGlnLeuAlaThrArgGlnLysAlaGlnGlnGlnLysSerSerAl 1064
2787 AGATGGCGCGCGCCGCTTCCGCGCGCTTCGCGCGCTTCCCTATTATCCG 2836
1064 aGlnLeuAlaArgArgHisGlnLysGlnArgGlnAlaLysLeuSerAl 1081
2837 TTACACCGCAACTTCGCTAGAACCTTTCACACGCTGACGCGTAAC 2886
1081 laLysGlnLysValGlnLysAlaGlnArgGlnAlaGlnAlaLeuAlaValArg 1097
2887 GCGCAATTCAGCGCTCAGGGAACATTCGCTTATGTCGCAACTCTT.CG 2935
1098 ArgLysAlaGlnAlaGlnLysAlaLysArgGlnAlaAlaGlnLeuAlaArg 1114
2936 GCTACCGGACGACAAATTTAGCT..... 2960
1114 gArgHisGlnLysGlnArgGlnAlaAlaGlnLeuSerAlaLysGlnArg 1131
2961GGCGGAAAGTTCGGAAGCATTACACTTGGCGGCTCAACAATACCGG 3008
1131 aGlyGlnGlnGlnLysArgGlnThrAlaGlnSerGlnProGlnArgArg 1147
3009 CACGCAACCTGCAAGCCTCGAACCAATTCAGCGGTAGTGGAAG..... 3050
1148 LysArgAlaGlnAlaProGlnAspTyrMetAlaAlaSerGlnAspArgTr 1164
3050 3050
1164 OlyArgArgGlyHisArgSerValGlnGlnAsnAsnValGlnIleAlaG 1181
3051AAAAGACACAA..... 3062
1181 lAlaGlnAlaGlnLeuAlaArgArgGlnGlnGlnLysAlaAla 1197
3062 3062
1198 GlnLeuLeuAlaLysGlnAlaArgAlaGlnAlaGlnLysGlnAlaLale 1214
3063 .ACCGCTGTCCGAAACCTTAATTTACACCTTCGCAAGCAACGCGCAT 3111
1214 uAlaAlaLysArgLysAlaGlnAlaGlnGlnAlaLysArg.GlnAlaAla 1230
3112 GCGCGCGCTGGCGGTTACCAACTCATCCGCAAGACGCGAGTTCGCGCT 3161
1231 GlnLeuAlaHisArgGlnGlnAlaGlnLysArgLysAlaAlaGlnLeuSerAl 1247

A:Reference number: A81000; MUID:20175755
 A:Accession: C81169
 A:Molecule type: DNA
 A:Residues: 1-1815 <TEP>
 A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AP41117.1; PID:g722592
 A:Experimental source: serogroup B, strain MC58
 R:Lehmann, H.; Poulsen, K.; Mogens, K.
 M01. Microbiol. 15, 495-506, 1995
 A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria*
 A:Reference number: S61314; MUID:95502961
 A:Accession: S61326
 A>Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 53-548 <LOW>
 A:Cross-references: EMBL:X82477; NID:g732856; PIDN:CA57860.1; PID:g732857
 C:Genetics:
 A:Gene: NMB0700
 C:Superfamily: Iga-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1491.50 Length: 2000
 Ratio: 1.637 Gaps: 56
 Percent Similarity: 45.550 Percent Identity: 24.550

alignment_block:

US-09-303-518d-649 x C81169

Align seg 1/1 to: C81169 from: 1 to: 1815

```

49 AAGACGGCGCATCGCTTCGCGCTGCTTACTTACGCAATAGCCCTGTC 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATTTCTCCCAAGCCTGGCGGACACACTTATTTCGGCATCA 148
   ::::: ::::: ||| ::::: ||||| ::::: :::::
   18 ATyrAlaLeuThrProTyrSerGluAlaIleValAlaIleArgAspValA 35
149 ACTACCAATACTATCGCGACTTGGCCGAAATTAAGGCAAGTTGCACTG 198
   ::|||::| ::|||::| ::|||::| ::|||::| ::|||::|
   35 spyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
199 GGGCGCAATATGATGAGGTGTACAAACAAAGGCGGCTGGTGGCGCAA 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   52 GluAlaIleThrAspLeuSerValLysAsnLysGlnLysAsnIleGlyAs 68
249 ATCAATGACAAAGCCCGATGATGATTTCTGTGGTGGCGGTAAAC. 297
   ::::: ::::: ||||| ||||| ||||| ||||| |||||
   68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaIleAspValAsnA 85
298 ..GGGTGGCGCATTTGGTGGCGCATCAATATGATGAGCGTGCACAT 345
   ::::: ::::: ||||| ||||| ||||| ||||| |||||
   85 rGATrThrLeuThrValIleAspProGlnTyrAlaValIleSerValLysHis 101
346 AACGGCGCTATTAACAAGCTGATTTGGTGGCGGAGGAGAAATCCCGA 395
   ||| ::::: ||||| ::::: ||||| ::::: |||||
   102 ValLysGlyAspGlnIleSerTyrTyrGlnHisHisAsnLysIleAsn 118
396 T.....CAACATGTTTACTTAAATGTAAGGAAAGGAAATTAATATA 439
   ::::: ::::: ||||| ||||| ||||| ||||| |||||
   118 pValSerAsnAspGlnAsnGlnTyrArgSerValAlaGlnAsnAspTyrC 135
440 AAGCAGGACTAAAGGCATCTTATGCGGC.....GAT 474
   ::::: ::::: ||||| ||||| ||||| ||||| |||||
   135 LurProAsnLysAsnTrpHis...HisLysAsnGlnLysArgLeuGluAsp 150
475 TATCATATGCCCGTGTTCATTAATTTGTCAGATGCAAGAACCTGTGTA 524
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   151 TyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaIleProIleAl 167
525 AATGACAGGTATATATGATGGG...CGGAATATATGATCAAAATTAAT 571
   ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

167 aProThrSerAlaGlyLysValGluThrTyrLysAspLysAsnArgP 184
572 ACCCTGACCGGTTCGATTTGGGCGAGCGACGCAATATTTGGCATCTGAT 621
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
   184 heserGluPheValArgValGlyAlaGlyThrGlnPhe..... 196
622 GAAGATGAGCCCAATACCGCGAAAGTTCATATCATATGCAAGTGGCGTA 671
   ||| ||||| ||||| ::::: ||||| ||||| |||||
   197 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyr 211
672 TTCTGGCTGCTGGTGGCATACCTTT.....GCACAAA 706
   | ::::: ||||| ::::: ||||| ::::: |||||
   211 rArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerA 228
707 ATGATCATGAGTGGTGGCACATCACTTAGTAGTGAATAAATTAACAT 756
   || ::::: ||||| ::::: ||||| ::::: |||||
   228 snLeuAsnGlnLysGlyLeuIleGlyPheGlyAspAsnSerLysHis 244
757 AGCCCA.....TAGGTTTTTACCAACAG 782
   ||||| ||||| ||||| ||||| ||||| |||||
   245 SerProGlnLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyr 261
783 AGGCTCATTTGGCGACATGGCGCACACCAATGTTATTCATGATGCCCAA 832
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
   261 rAlaValIleGlyAspSerGlySerProLeuPheAlaTyrAspLysGln 278
833 AGCAAAAGTGTAAATTAATGGGGTATTCACAAACGGCAACCCCTATTA 882
   ::::: ||||| ::::: ||||| ::::: |||||
   278 LysrArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr... 293
883 GGAANAAGCAATGGCTTCACAGCTGGTGTGAAGATTG...TTCTATGA 929
   ::::: ||||| ::::: ||||| ::::: |||||
   294 ...GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLys 305
930 TGAATCTTTGCTGAGATACCATTCAGTATTCACCAACAGCTGAA 979
   ::::: ||||| ::::: ||||| ::::: |||||
   305 LysGluPheAlaAsp.....GluIleLysGln 315
980 ATGGGAATACTCTTTTAACGACATTAATGCGACAGGAAATCAAT 1029
   ::::: ||||| ||||| ||||| |||||
   315 rG.....AspAsnAlaGlyThrIleLysGlyAsn 324
1030 GCCAAACATGAACACATTCCTGCTGCTAATGATTAACACACACACCT 1079
   ::::: ||||| ::::: ||||| ::::: |||||
   325 GluGlnHisHisTrpLysThr.....ThrGlyThr... 334
1080 TCAATTTGTTAATGTTCTTATCCGACACAGCAAGAACCTGTATTC 1129
   ::::: ||||| ||||| ||||| |||||
   335 .....AsnSerHisIleGlySerThr..... 341
1130 ATGCTGAGGTGGTGTCAACAGTTATGACCCAGACTGAATATGCGAA 1179
   ||||| ||||| ||||| ||||| ||||| |||||
   342 ..AlaValArgLeuAlaAsnAsnGlnLysArgP...AlaAsnAsnGln 356
1180 AATATTTCTTATTTGACGAGGAAAGCGCAATGATGATCTTCCAGCAA 1229
   ||||| ||||| ||||| ||||| ||||| |||||
   357 AsnValThrPhe.....GluAsnAsnGlyThrLeuValIleAsnGln 371
1230 CATCAATGAAGTGTGAGAGATTAATTTCAAGAGATTTTACGATC 1279
   ||||| ||||| ||||| ||||| ||||| |||||
   371 nIleAsnGlnIleValaGlyGlyLeuPhePheLysGlyAspTyrThrVal 388
1280 CGCCTGAAATATACGAA...ACTTGGCAAGCGCGGCTTCATATCACT 1326
   ::::: ||||| ||||| ||||| ||||| |||||
   388 ysgLysIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAla 404
1327 GAACACAGTACCGCTTACTTGAAGAGTAAGCGGCTGCAAGACCGCGCT 1376
   ::::: ||||| ||||| ||||| ||||| |||||
   405 AspLysLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLe 421
1377 GTCCAAATATGCGCAAGGCGCTGACAGTTCACAGCAAGGCAAGCAAC 1426
   ||||| ||||| ||||| ||||| ||||| |||||
   421 uAlaLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnG 438

```


[illegible]

```

seq_name: p1r2:C41859
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C|Species: Haemophilus influenzae
A|Variety: strain HK613
C|date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C|Accession: C41859
R:Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A|Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A|Reference number: A41859; MUID:92234949
A|Accession: C41859
A|Status: preliminary; not compared with conceptual translation
A|Molecule type: nucleic acid
A:Residues: 1-1849 <POU>
A:Experimental source: strain HK613
A>Note: Sequence extracted from NCBI backbone (NCBIP:97285)
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 1487.50 Length: 1917
Ratio: 1.492 Gaps: 59
Percent Similarity: 52.008 Percent Identity: 24.935

alignment_block:
US-09-303-51BD-649 x C41859 ..

Align seg 1/1 to: C41859 from: 1 to: 1849

64 CGCTTCGCGCTGCTTACTTAGCCATATGCCTGCTGGCATTTCTTC 113
:::||||: :::::::::::::::::::: ||
5 LysPheLysLeuAsnPhelIleAlaLeuThValAlaIylAlaLeuHrp 21
:::||||: :::::::::::::::::::: ||
114 CCAAGCCTGGGCGGACACACTTATTGCGCATCACTACCAATPATATC 163
||:::||||: :::::::::::::::::::: ||
21 oTYrThGlAlAlaIAlaLeuValArgAspAspValAsPTyGLIlIeph 38
||:::||||: :::::::::::::::::::: ||
164 GCGACTTTGCCGAATAAAGCAAGTTTGCAGTGGGCGGAAAGATATT 213
||:::||||: :::::::::::::::::::: ||
38 rGaSpHealaglLuAsnLysgLyLysPheSerValglYAlaThrAsnVal 54
||:::||||: :::::::::::::::::::: ||
214 GAGTTTACAACAATAAAGGGAGTTGGTCGCAATCAATGACAAAGC 263
||:::||||: :::::::::::::::::::: ||
55 GluValArgAspLysAsnGlnSerLeuGlSerAlaLeuProAsnGI 71
||:::||||: :::::::::::::::::::: ||
264 C...CCGATGATGATTTTCTGTGTGTG...CGTAACGGCGTGGCG 307
:::||||: :::::::::::::::::::: ||
71 yIlePrometIleAspPheSerValValAspValAspLysArgIleAlar 88
||:::||||: :::::::::::::::::::: ||
308 CATTTGGTGGCGATCATATATTGTGAGCGTGGCAACT..AACGGGGGC 354
||:::||||: :::::::::::::::::::: ||
88 hrLeuValAsnProGlnTrValValglValLysHisValSerAsngly 104
||:::||||: :::::::::::::::::::: ||
355 TATAACAAGCTTGATTTTGGT.....GCGGAAGGAGAAGAA 389
||:::||||: :::::::::::::::::::: ||
390 TCOCGATCACATCGTTTACT.....TATAAAATTG 421
||:::||||: :::::::::::::::::::: ||
121 nAlaLysSerHisArGAspValSerSerGluGluAsnArgTyryTyrv 138
||:::||||: :::::::::::::::::::: ||
422 TGAAACGAATATATAT.....AAA 441
||:::||||: :::::::::::::::::::: ||
138 aLGluLysAsnAsnPherProThrGluAsnValThrSerPheThrThlys 154
||:::||||: :::::::::::::::::::: ||
442 GCAGGAGCTAAAGCCATCTTATGGCGGCAATTATCATATCCGCGTTT 491
||:::||||: :::::::::::::::::::: ||
155 GIuGIuGlnAspAlaGlnLysArgIrgrGIuAspTyryTyrmelProArg 171
||:::||||: :::::::::::::::::::: ||
492 GCATAAATTTGTGCACAGATGCAGAACTGTTGAAATGACCAGTTATATG 541

```

171 uasplyseheValThrGluValAlaProIleGluAlaSerThrAlaAsn 188
542 ATGGGCGG...AAATATATGATCAAAATATATACCTGACCGTGTCT 588
188 snAsnLysGlyIuTyraSnAsnSerAspLysTrProAlaPheValArg 204
589 ATTGGGCGGAGGAGCAATAT... 609
205 LeuGlySerGlySerGlnPheIleTyrlsGlySerArgTyGlnLe 221
610 .TGCGCATCTGATGAAGATGAGCCCAATACCGCAAGTTCATATCAT. 657
221 uileuThrGluLysAspLysGlnLysAsnLeuArgAsnTrpAspV 238
658ATGCAAGTGGCATCTGCTGCTGCT 684
238 alGlyLysAsnLeuGluLeuValGlyAsnAlaIleTyrlsGlyIle 254
685 GGTGGCAATACCTTTCACAAATGATGAGTGGTGCACAGTCAACT 734
255 AlaGlyThrProTyrlsValAsnHisGluAsnAsnGlyLeuIleGlyPh 271
735 AGTATGATGAAAAATTAACATAGC...CCATATGTTTATTCACACAG 781
271 eGlyAsnSerLysGlnGlnHisSerAspProLysGlyIleLeuSerGlnA 288
782 GA.....GGCTCATTTGGCGACAGTGGCTCACCAATGTTT 816
288 sPrProLeuThrAsnTyrlaValLeuGlyAspSerGlySerProLeuPhe 304
817 ATCTATGATGCCCCAAAGCAAAAGGTATTAATGCGGTATTCGAAC 866
305 ValTyraPheArgGluLysGlyLysTrPheLeuGlySerTyraPhe 321
867 GGGCAACCCCTATATGAGAAAGCAATGCTTCACGCTGTTGCTGAAG 916
321 eTrpAlaGlyTyraSnLysSerTrpGlnGluTrpAsnIleTyrlsH 338
917 ATTGGTCTATGATGAATCTTTCGTGACATCCCATCATGATTTTAC 966
338 IsGlnPheAlaGluLysIleTyrls...GlnGlnTyrlsSerAlaGlySer 352
967 GAACCACTCAAAATGGCAATACTCTTTACAGACATATATATGCGAC 1016
353 LeuThrGlySerAsnThrGlnTyrlsTrp.....GlnAlaTh 365
1017 AGGAAAAATCAATGCCAAACATGATCTGCTCAATATATATTA 1066
365 rGly.....S 367
1067 AAACACGAAACCGT.....CAATTGTTTAATGTTTCTTATATCC 1104
367 eTrhSerThrIleThrGlyGlyLeuProLeuSerValaPheLeuThr 383
1105 GAGACAGCAGAGAACCTGTTTATCATGCTGAGAGTGTGTCAACAGTA 1154
384 AspLysLysAspLysPro..... 389
1155 TCGACCCAGACTGAATATGAGAAAAATATTCCTTATTCAGCAAGAA 1204
390AsnHisGlyLysSerIleThrLeu.....LysGlyLys 400
1205 AAGCGCAATGATACCTTACAGCAACATCAATCAAGTGTGAGAGATTA 1254
400 eTrGlyThrIleThrLeuAsnAsnHisIleAspGlnIleAlaGlyLeu 416
1255 TATTCGAAGAGATTTTAACGTC...TGCGTGAATATACGAACCTTG 1301
417 PhePheLeuLysPtyrGluValLysGlyThrSerAspSerThrThrTr 433
1302 GCAAGGCGGCGGCTTCATATCAGTGAAGACAGTACCGTTACTTGAAG 1351
.....
433 pLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysV 450
1352 TAAACGGCGTGGCAAAAGCCGCTGTCCAAAATGGCAAGCAACGCTG 1401
450 alHisAsnProLysTyraPheArgLeuAlaLysIleGlyLysGlyThrLeu 466
1402 CACGTTCAACCCAAAGGGGAAAAACCAAGCTGATCAGTACGCTGGGAGAG 1451
467 ValValGlnIuGlyLysGlyLysAsnGlnGlyLeuLeuLysValGlyAspI 483
1452 TACAGTCATTTTGGATCAGCAGCAGACAGATTAAGCAAAAAACAGCT 1501
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaP 500
1502 TTAGTAAATCCGCTGTGTCAGCGAGGCGTACGCTGCAACTGAAATCC 1551
500 heSerGlnValGlyIleValaSerGlyArgSerThrLeuValLeuAsnAsp 516
1552 GATATACACTTCAACCCGCAAACTCTATTTCGCTTTCGCGCGAGC 1601
517 AspLysGlnValaAspProAsnSerIleTyrlsPheArgGlyGlyArg 533
1602 TTTGATTTAAACGGGCAATTCGCTTTCGTCACCGTATTCAAAATACG 1651
533 gLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleA 550
1652 ATGAAGGGCGGATGTGTCAACCAATCAAGCAAAAGATCCAGCTT 1701
550 sPAspGlyAlaArgValaValaSnHisAsnMetThrAsnThrSerAsnIle 566
1702 ACCATTACAGCG.....ATAACATATTTGCTTACACCGG 1736
567 ThrIleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTy 583
1737 CAAT..... 1740
583 rAsnIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIleP 600
1740 1740
600 rOTyrlsArgGlnLeuTyrlsPheAsnGlnAspAsnArgSerTyrlsThrLeu 616
1740 1740
617 LysLysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyLysE 633
1740 1740
633 rAsnGluAsnTrpLeuTyrlsMetGlyArgThrSerAspGluAlaLysArg 650
1741AAACACAGCTGTGATAGCAAAAAAGAAATTCGCTTACCAACGCTTG 1785
650 snValMetAsnHisIleAsnAsnGlnIuTyrlsMetAsnGlyPheAsnGlyTyrls 666
1786 TTTGGCGAGAAATACAGCAAAACG...AACGGCGGCTCAACCTTGT 1832
667 PheGlyGlnGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValaTh 683
1833 TTAACACCCCGCGGAGAGAGCAACCCGCTGCTGCTTTCGCGGAGCA 1882
683 rPheAsnGlyLysSerAspLeuAsnArgPheLeuLeuThrGlyGlyThrA 700
1883 ATTTAAACGGCAACATCAGCAAAACCAAGCGCAAACTGTTTTCAGCGCG 1932
700 snLeuAsnGlyAspLeuAsnValaGluLysGlyThrLeuPheLeuSerGly 716
1933 AGACCAACCGGAGCGCTACATCATTTTAAACGACCAATGGTGGCAAA 1982
717 ArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLys 733
1983 AGAGGCG.....ATTCTCGGCGGGAATCTGTGGGACCAACGCTGGA 2026
733 sAspProHisPheThrGluAsnAsnGluValaValaGluAspAspTrpI 750

C:\Superfamily: IGA-specific metalloendopeptidase

```
seq_name: pIrf2:I54632  
seq documentation block:  
tsh protein - Escherichia coli  
C:Species: Escherichia coli  
C>Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Dec-2000  
C:Accession: I54632  
R:Providence, D.L.; Curtiss, R.  
A>Title: Isolation and characterization of a gene involved in hemagglutination by an infect. Immun. 62, 1369-1380, 1994  
A:Reference number: I54632; MIMD:94178945  
A:Accession: I54632  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
M:Molecule type: DNA  
A:Residues: 1-1377 <RES>  
A:Cross-references: GB:I27423; NID:g469235; PIDN:AAZ4698.1; PID:g469236
```

alignment_scores:

Quality: 1111.50 Length: 1573
Ratio: 1.388 Gaps: 62
Percent Similarity: 50.922 Percent Identity: 25.366

alignment_block:

US-09-303-518D-649 x 154632 ..

Align seg 1/1 to: 154632 from: 1 to: 1377

```

90 ATCCCTGCTGCTGGCATTCTTCCCAAGCCTGGGGGACACACTTAT 139
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
42 ILeProValIeu...PheSerAlaGlySerLeuAlaGlyThr.ValAsnA 57
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
140 TCGGCATCACTACCAATACCTATCGGACTTGGCGAATAAAGGCAAG 189
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
57 snGluLeuGlyTYrGlnLeuPheArgAspPheAlaGluAsnLysGlyMet 73
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
190 TTTCACCTCGGGCGGAGAAATATTGAGTTTACACAAAAAGGCAAGT 239
   ||| ||||| ::::::::::: :::::::::::
74 PheArgProGlyAlaThrAsnIleAlaIleTYrAsnLysGlnGlyLuph 90
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
240 GGTGGCAATTCATGACAAAGCCCGATGATTGATTTTCTGTGTG- 288
   ::::::::::: ::::::::::: ||| ||||| :::::::::::
90 eValGly...ThrLeuAspLysAlaAlaMetProAspPheSerAlaValA 106
   ::::::::::: ::::::::::: ||| ||||| :::::::::::
289 .TCGCGTAACGGCGTGGCGCATTTGGTGGCATCAATATATTGTGAGC 336
   ::::::::::: ||||| ::::::::::: :::::::::::
106 spSerGluIleGlyValAlaIleThrLeuIleAsnProGlnTYrIleAlaSer 122
   ::::::::::: ||||| ::::::::::: :::::::::::
337 GTGCACATACAGCGCGCTATACACAGCTGATTGTGTGGCGAAGGAG 386
   ||| ||||| ::::::::::: :::::::::::
123 ValLysHisAsnGlyTYrThrAsnValSerPheGly...AspGlyG 138
   ||| ||||| ::::::::::: :::::::::::
387 AATCCGATCAACATGCTTTACTATTAATTTGTAACGGAATATT 436
   ||| ::::::::::: ||||| :::::::::::
138 uAsn.....ArgTYrAsnIleValAlaSprArgAsnAsnA 149
   ::::::::::: ::::::::::: ||||| :::::::::::
437 ATAAAGCAGGAGCTAAAGGCCATTCATGCGCGCATATCATATGCCG 486
   ::::::::::: ::::::::::: ||||| :::::::::::
149 LaProSer.....LeuAspPheHisAlaPro 157
   ::::::::::: ::::::::::: ||||| :::::::::::
487 CGTTGCTAAATTTGTCACAGATGCAAGCCTGTGAAATGACCACTTA 536
   ||||| ::::::::::: :::::::::::
158 ArgLeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaG 174
   ||| ::::::::::: :::::::::::
537 T...ATGATGGGGGGAATATATTCGATCAAAATTAATTCACCTGAC 583
   ::::::::::: ||||| :::::::::::
174 nGlyAlaValAlaGlyAlaTYrLeuAspLysGluArgTYrProValPheT 191
   ::::::::::: ||||| :::::::::::
584 TTCGATTTGGGGGAGCAGCAATTTGGCATCTGATGAGAGAGAGGCC 633
   ||| ::::::::::: :::::::::::
191 yArgLeuGlySerGlyThrGlnTYrIleLysAspSerAsnGlyGlnLeu 207
   ::::::::::: ||||| :::::::::::
634 AATAACCGGAAATTCATATCATATTCGAAGTCGATTTGGCTCGT 683
   ::::::::::: ::::::::::: ||||| :::::::::::
208 Thr.....GlnMetGlyAlaTYrSerTrpLeuTh 218
   ::::::::::: ::::::::::: ||||| :::::::::::
684 TGGTGGCAATACCTTGCA.....CAAATGA..... 711
   ::::::::::: ::::::::::: ||||| :::::::::::
218 rGlyGlyThrValGlySerLeuSerSerTYrGlnAsnGlyLumetLies 235
   ::::::::::: ::::::::::: ||||| :::::::::::
712 ..TCGAGTGGGCGACAGTCACACTTAGTAGTGAATAAATTAACATAGC 759
   ::::::::::: ||| :::::::::::
235 eTrhSerSerGlyLeuVal.....PheAspTYrLys 245
   ::::::::::: ::::::::::: ||||| :::::::::::
760 CCATATGCTTTTATACCAACAGAGGCTCATTTGGCGACAGTGGCTCAC 809
   ||| ::::::::::: ||||| :::::::::::
246 LeuAsnGlyAlaMetProLethTYrGlyAlaGlyAspSerGlySerTr 262
   ::::::::::: ::::::::::: ||||| :::::::::::
810 AATGTTTATCTATGATGCCCAAGCAAAAGTGTTAATTAATGGGGTAT 859
   ::::::::::: ::::::::::: ||||| :::::::::::
262 oLeuPheAlaPheAspThrValGlnAsnLysTrpAlleuValGlyValL 279

```

```

860 TCGAAAGCGGCAACCCCTATATAGCAAAAAAGCAATGGCTTCACAGCTG 909
   || ::::::::::: ||::: :::::::::::
279 eutrhAlaGlyAsnGlyAlaGlyValrGlyAsnAsnTrpAlaValIle 295
   || ::::::::::: :::::::::::
910 CGTAAAGATTGGTCTATGATGAATCTTTCCTGAGATACCAATTCAGT 959
   || ::::::::::: :::::::::::
296 ProLeuAspPheIleGlyGlnLysPheAsnGlnLysPasnAlaProVa 312
   ::::::::::: :::::::::::
960 ATTCTACGAAACCGTCAAAATGGG.....AAATACTCTTTTACGACG 1003
   ::::::::::: :::::::::::
312 lThrPheArgThrSerGlnGlyAlaLeuGlnTrpSerPheAsn...S 328
   ::::::::::: :::::::::::
1004 ATAAATATGCGACAGGAAATAATCAATGCCAAACATGAACAATTCCTG 1053
   ::::::::::: ||| :::::::::::
328 eSerThrGlyAlaGly..... 333
   ::::::::::: :::::::::::
1054 CCTAATAGATTAAAAACAGAACCGTTCATATGTTTATGTTCTTATAC 1103
   ::::::::::: :::::::::::
334 .....AlaLeuTh 336
   ::::::::::: :::::::::::
1104 CGAGCAGCAGCAAGACCTGTTATACGTCGAGGTGGTGCACAGTT 1153
   ::::::::::: ||| :::::::::::
336 rGlnGlyThrThrTYrAlaMetHisGlyGlnGlnLysAsnSP... 351
   ::::::::::: ||| :::::::::::
1154 ATCGAACCCAGACTGAATATGAGGAAATATTCTTATTGACGACGAGA 1203
   ||||| ||||| :::::::::::
352 .....LeuAsnAlaGlyLysAsnLeuIlePheGlnGlyLyn... 363
   ::::::::::: :::::::::::
1204 AAAGCGAATTGATCTTACCAAGCAACATCAATCAAGTGTGGCGAATT 1253
   ::::::::::: ||| :::::::::::
364 AsnGlyGlnIleAsnLeuLysAspSerValSerGlnGlyAlaGlySerLe 380
   ::::::::::: ||| :::::::::::
1254 ATATTTCACAGAGATTATGAGCTCTCGCTGAAATATACGAACTTGGC 1303
   ||| ::::::::::: ||||| :::::::::::
380 uThrPheArgAspAspTYrThrValThrThrSerAsnGlySerTrpTrpT 397
   ::::::::::: ::::::::::: ||||| :::::::::::
1304 AAGGCGCGGGCTTCATATCACTGAGACAGTACCGTTACTTGGAAACTA 1353
   ||||| ::::::::::: :::::::::::
397 hrcGlyAlaGlyIleValAlaAspAsnGlyValSerValAsnTrpIleVal 413
   ::::::::::: ||||| :::::::::::
1354 AAGCGGTGGCAAGCAGCGCGCTGTCAAATTCGCAAGGACCACTGCA 1403
   ||||| ::::::::::: ||||| :::::::::::
414 AsnGlyValLysGlyAspAsnLeuHisLysIleGlyGlyGlyThrLeuTh 430
   ::::::::::: ||||| :::::::::::
1404 CGTTCAAGCCAAAGGGGAAACCAAGCTCGATCAAGCTGGCGACGGTA 1453
   ||||| ::::::::::: ||||| :::::::::::
430 rValGlnGlyThrGlyIleAsnGlnGlyLeuLysValGlyAspGlyL 447
   ::::::::::: ||| :::::::::::
1454 CAGTCATTTTGATCGACAGCGCAGATTAAGGCAAAAAACAAGCTTT 1503
   ||| ::::::::::: ||||| :::::::::::
447 yValAlaLeuAsnGlnGlnAlaAspAsnLysGlyGlnValGlnAlaPhe 463
   ::::::::::: ||||| :::::::::::
1504 AGTGAATGCGCTGTGTCAGCGGAGGAGTACGCTCAACTGAATGCCA 1553
   ||| ::::::::::: ||||| :::::::::::
464 SerSerValAsnIleAlaSerGlyArgProThrValLeuThrAspL 480
   ::::::::::: ||||| :::::::::::
1554 TAATCACTTACACCCGACAACTATTTCCGCTTCCGCGGAGAGCTT 1603
   ::::::::::: ||||| :::::::::::
480 uArgGlnValAsnProAspThrValSerTrpGlyTYrArgGlyGlyThL 497
   ::::::::::: ||||| :::::::::::
1604 TGGATTTAACGGGCACTGCTTTCCTTCCACCGTATCAAAATACCGAT 1653
   ||||| ::::::::::: ||||| :::::::::::
497 euAspValAsnGlyAsnSerLeuThrPheHisGlnLeuLysAlaAlaAsp 513
   ::::::::::: ||||| :::::::::::
1654 GAAGGGCGATGATTTCAACACCAATCAAGACAAGATCCACCGTTAC 1703
   ||||| ::::::::::: ||||| :::::::::::
514 TYrGlyAlaValAlaLeuAlaAsn...AsnValAspLysArgAlaThrIleTh 529
   ::::::::::: ||||| :::::::::::
1704 CATTACAGGCATTAACATATTTGCTACACCGGCAATTAACAACAGCTGG 1753
   ::::::::::: ||| :::::::::::
529 rLeu.....AspTYrAlaLeuArgAlaAsp..... 537

```


1754 ATAGCAAAAAGAAATTCCTACACGGTTGG..... 1785
538LysValAlaLeuAsnGlyTrpSerGlySerGlyLysGly 550
1785 1785
551 ThrAlaGlyAsnLeuTyrLysTyrAsnAsnProTyrThrAsnThrThrAs 567
1786TTGGCGGAGAAAGATAGAC..... 1806
567 pTyrPheLeuLeuLysGlnSerThrTyrGlyTyrPheProThrAspGlns 584
1806 1806
584 eSerAsnAlaThrTrpGluPheValGlyHisSerGlnGlyAspAlaGln 600
1807AA 1808
601 LysLeuValAlaAspArgPheAsnThrAlaGlyTyrLeuPheHisGlyGly 617
1809 AACGACGGGGCGCTCAACCTGTTTACAG...CCC GCCGAGAGAC 1855
617 nLeuLysGlyAsnLeuAsnValAspAsnArgLeuProGluGlyValThrG 634
1856 GCAACCTGCTGCTTCCGGGAGAAATTAACGGCAATCAGCA 1905
634 LysAlaLeuValMetAspGlyAlaAlaAspIleSerGlyThrPheThrGln 650
1906 ACAACGGCAAACTGTTTTCAGCGGAGACCAACCGCAGCCTACAA 1955
651 GluAsnGlyArgLeuThrLeuGlnGlyHisProValIleHisAlaTyrAs 667
1956 TCATTTAAAC.....GACCAT..... 1971
667 nThrGlnSerValAlaAspLysLeuAlaIleSerGlyAspHisSerValL 684
1972TGTCGCAAAAAGAGCGCATTCCTCGCGGCA 2004
684 euthrGlnProThrSerPheSerGlnGln..... 693
2005 ATCGTGTGGACACGACTGATCAACCGCACATTTAAAGCGGAAACTT 2054
694AspTrpGluAsnArgSerPheThrPheAspArgLe 705
2055 CCAATTAAGCGGACAGCGCGGTGTTCCGCAATGTTGCCAAAGTGA 2104
705 uSerLeuLysAsnThrAspPheGlyArgAsnAlaThr..... 719
2105 AAGCGATGGCATTTGACCAATCAGCGCCAAAGCA...GTTTGGGTGC 2151
720LeuAsnThrThrIleGlnAlaAspAsnSerSerVal 731
2152 GCACCGCATCAAGCCACACATCTGTACAGCTTCGACTGAGCGGTCT 2201
732 ThrLeuGlyAspSerArgValPheIleAspLysAsnAspGlyGlnGlyTh 748
2202 GACAATTTGTGTGCAAAA.....ACCATTAACGACGATTAAGTGA 2242
748 rAlaPheThrLeuGlnGlyThrSerValAlaThrLysAsp..... 762
2243 TTGCTTCACTAGACGACATCAGCGCATGTCGATCTCGCGCAT 2292
763AlaAspLysSerValPheAsnGlyThrValAsnLeuAspAsn 776
2293 CACGCTCATTTAAATCTCACAGGGCTTGCACACTCAAGCGCAATCTTAG 2342
777 GlnSerValLeuAsnIleAsnAspIle.....PheAsnGlyGlyIleGly 791
2343 TGCAATGTGGCATACGTTATACAGTCACCAACGCAACCCCAAGC 2392
791 nAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804
2393 GCAACCTTAGCTCGTGGGCAATGCCAACCAATTTAATCAAGCACA 2442
805ValLeuGlyAsnSerThrLeuThrSerThrAlaLeuAsn 817
2443 TTAACGGCAACATCGGCTTCGGCAATGCTTCATT..... 2481
818 LeuAsnLysGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAspGly 834
2482AATCTAAGCAGCACCGCCGACAA..... 2505
834 yProValAsnIleSerAspAlaAlaLeuSerLeuAsnSerArgProAspG 851
2506AACGCACTGTGACG 2520
851 LuValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerThrAsn 867
2521 CTTCCTCCGC...AACGCTAAGCAACGTAAGC...CATTCGCACTCA 2564
868 LeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeuSe 884
2565 CGGTATGTCTCCCTAGCCGATAG..... 2589
884 rGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGlyGln 901
2590CGAGTA 2595
901 LysLeuSerProAspPheThrLeuGlnAsnGlnMetLeuTyrSerLeu 917
2596 TTCATTTTGAAGACAGCCGCTTACCGGACAAATCAGCGGCGCAAGA 2645
918 PheAsnGlyTyrArgAsnIleTrpSerGlySerLeuAsn...AlaProAs 933
2646 TACGCAATTAACCTTAAAGACAGCAATGACGCTGCCGTCAGGACGG 2695
933 PalatThrValSerMetThrAspThrGlnTrpSerMetAsnGlyAsnSer 950
2696 AATTGAAGCAATTAACCTTGAACAAGCCACCATCACTCAATTCGCC 2745
950 hValaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsnGlyGly 966
2746 TATGCCACGATGGCGGAGGCGCAACCGCAGTCGACAGATGGCC 2795
966 966
2796 GCGCGCGCTTGGCGCGCTTCGCCCTATATCCGTTACACGCC 2845
966 966
2846 CAACTTCGTAGATCCCGTTTCAACAGCTGACGTAACGCAATTTG 2895
967ThrSerProPheThrThrLeuThrThrAspAsnLeuAsp 979
2896 AACGCTCAGGACATTCCTTATGTGCGAACTTCGCTACCGCAG 2945
980 AlaValGlnSerAlaPheValMetArgThrAspLeu.....AsnLysAl 994
2946 CGACAATTTGAAGCTGGCGGAAAGTTCGAAAGGACACTTACACCTTGGCG 2995
994 aAspLysLeuValIleAsnLysSerAla..... 1003
2996 TCACATATCCGGCAACGACACTCAAGCCTCGAACATTTGACGATAGT 3045
1003 1003
3046 GAAGGAAAGACACAAACCGCTGTCCGAAACCTTATATTCAACCTGCA 3095
1004 ThrGlyHisAspAsn.....SerIleTrpValAsnPhe..... 1014
3096 AAACGAACAGCTGATCGCGCGGTGCGCTTACCAACTCATTCGCAAG 3145
1014 1014
3146 ACGCGAGTTCGCTTCATATCCGTTCAAGACAAAGAGCTTTCGAC 3195

```

1014 ..... 1014
3196 AACTCGGCAAGGCAAGCAACCAAAAAAGGAGGAAAAAGCAACGGCGCA 3245
1015 ..... 1022
3246 AAGCCTTGACGGCGTGTGATGCGCGGCGCGATCCGTCGAAAAGACAG 3295
1023 ..... 1033
3296 AAGCGTGGCGAAGCGCGCGCGCGCGCGGAGGAAAAATGTCGGC... 3342
1033 LuAlaThraLaAspaNLeuPheAArgAlaSerThrArgValAlaGlyPhe 1049
3343 ..... ATTATGACGGCGAGAGAAAGAAAAAGGAGTGA 3377
1050 SerAspValThrProIleLeuSerValArgLysGluAspGlyLysGlu 1066
3378 GGGGATTAAGACACCGCCTTGCGGAAACAGCGCGGAAACCGCGC 3427
1066 uTrpValLeuAspGlyTyrGlnValAlaArgAsnAspGlyGlnGlySA 1083
3428 CGGTACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCG 3477
1083 LaAlaAlaThr..... 1086
3478 CAACGTCAACCCCAACCGCAGCCCAACCGCAGCGACCTGATCAGCGC 3527
1086 ..... 1086
3528 TTATGCCAATAGCGGTTGATGATTTTCGCGACGCTCAACAGCGTT 3577
1087 ..... 1103
3578 TCGCGTACAGAGCAATAGACCGCGCTATTCCGAAAGACCGCGCCAC 3627
1103 snLysArgMetGlyAspLeuArgAspIleAsnGlyLys.....Ala 1116
3628 GCCGTTTGACA..... AGCGGATCCGGGAC..... 3654
1117 GlyThrTrpValArgLeuLeuAsnGlySerGlySerAlaAspGlyLys 1133
3655 ..... AACCAACATACCGTTCCGACAGATTCCGCGCTACCGCAACAAACCG 3703
1133 eThraspHisTyr.....T 1138
3704 ACCTCGGCCAAATCGCTATGCAGAAA.....AACCTCGGCAGCGGCGC 3747
1138 hLeuLeuGlnMetGlyAlaAspArgLysHisGluLeuGlySerMetAsp 1154
3748 GTC.....GGCATCCTGTTTGCACAAACCGACGCAAAAAACCTTGA 3791
1155 LeuPheThrGlyValMetAlaThrTyrThrAspHisAlaSerAlaAs 1171
3792 CGACGGCATCGCAACTCGGCAAGCGCTTCCGCGCGCGCTTTTCGGGC 3841
1171 pLeuTyrSerGlyLysThrLysSerTrpGlyGly..... 1183
3842 AATACGGCATCGACAGGTTTACATCGGCAATCAGCGCGCGCGGCTTT 3891
1184 ..... PheTyr.....AlaSerGlyLeuPhe 1190
3892 AGCAGCGCAGCCTTTCAGACGCGCATCGAGGCAAAATCCCGCGCGT 3941
1191 ArgSerGlyAlaTyrPheAspValIleAlaLysTyrIleHisAsn.... 1205
3942 GCTGATTACGGCATTCAGGCAAGATACCGCGCTTTCGGCGGATTCG 3991
1206 .....GluAsnLysTyrAspLeuAsnPheAlaGlyLag 1217
3992 GCATCAACCGCACATCGGCGCAACGCGTATTGTCCAAAACCGGAT 4041
1217 LysGlnAsnPheArgSerHisSerLeuTyrAlaGlyAlaGluValGly 1233

```

```

4042 TACCGCTACGA.....AACGTCAATATGCGCACCC... 4074
1234 TyrArgTyrHisLeuThrAspThrPheValGluProGlnAlaGlu 1250
4075 .....GGCCTTGATTCACCGCGCTACCGCGCGGCA 4105
1250 uValTrpGlyArgLeuGlnGlyGlnThrPheAsnTrpAsnAspGly 1267
4106 TTAAGCGATTTATTCATTCAACCGCGCGCAACATTTTCATCAGCGCT 4155
1267 et.....AspAlaSerMetArg.....ArgAsnSerValAsnPro 1278
4156 TATTGAGCGCTGCGCATACCGATCGCTTCGCGCAAAAGTCGCAACAG 4205
1279 Leu.....ValGlyArgThrGlyValValSerGlyLysThrPheSerG 1293
4206 CGTCAATACCGCGCTATTGCTCAG.....GATTTCCGCA 4240
1293 LysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyrGluPheAsp 1310
4241 AATCCCGCATGCGGAA 4257
1310 eThrAspSerAlaAsp 1315
seq_name: p1r2:S57664

```

```

seq_documentation block:
Iga-specific metalloendopeptidase homolog sepa precursor - Shigella flexneri
C:Species: Shigella flexneri
C:date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Dec-2000
C:accession: S57664; S69769; S69768
R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsoet, C.
submitted to the EMBL Data Library, February 1995
A:Description: Characterization of SepA, the major extracellular protein of Shigella
A:Reference number: S57664
A:Accession: S57664
A:Molecule type: DNA
A:Residues: 1-1366 <BEN>
A:Cross-references: EMBL:Z48219; NID:9886952; PIDN:CA88252.1; PID:9886953
R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsoet, C.
Mol. Microbiol. 17, 123-135, 1995
A:Title: SepA, the major extracellular protein of Shigella flexneri: autonomous secre
A:Reference number: S69768; MUID:96020667
A:Accession: S69769
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-66;1077-1366 <BEW>
A:Cross-references: EMBL:Z48219; NID:9886952; PIDN:CA88252.1; PID:9886953
A:Experimental source: strain M90T (serotype 5)
A:Accession: S69768
A:Molecule type: protein
A:Residues: 57-61, 'Q', '63-72', 'X', '544-557', 'X', '1057-1068 <BEF>
A:Note: 6-Glu, 14-Thr, 18-Gly and 26-Gln were also found
C:Genetics:
A:Gene: sepa
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: extracellular protein
F:1-56/Domain: signal sequence #status predicted <SIG>
F:57-1366/Product: Iga-specific metalloendopeptidase homolog sepa #status predicted <

```

```

alignment_scores:
Quality: 1001.50      Length: 1580
Ratio: 1.253          Gaps: 63
Percent similarity: 50.570      Percent identity: 25.886

```

```

alignment_block:
US-09-303-518D-649 x S57664 ..

```

```

Align seg 1/1 to: S57664 from: 1 to: 1366

```

```

22 ACAACCGAACAACACCGCAACCGCCGCAAAACCGGCGCATCGCTTCTC 71

```

```

111 .....|
27 ThrCysLysSerHisArg.....ArgLeuSe 35
72 GCGCTGCTACTATGACCATATGCGTGTGGATTCCTCCCAACCT 121
1 rArgArgValIleLeuThrSerValAlaAlaLeuSerLeuSerSerIleAt 35
122 GG.....GCGGACACACTTATTTGCGCATCACTACCAATACTAT 162
52 rPProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68
163 GCGGACTTTGGCGAAATAAAGCAAGTTGCACTCGGCGCGAAGATAT 212
69 ArgAspPheAlaGluAsnLysGlyGlnPheThrProGlyThrThrHisnIl 85
213 TGAGTTTACAACAAAAAGGGAGTGTGGCGCAATCAATCAACAAG 262
85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101
263 CCGCGATGATGATTTTCTGTGTGCGGTAC..... 297
101 lApProMetAlaAspPheSerSerAlaThrIleThrIleThrGlySerLeuPro 117
298 ...GGCTGGCGCATGTGTGGGATTCATATATATGTGAGCTGGCACA 344
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValValThrAlaLysHis 134
345 TAAACGGCGGTATACAAACGTTGATTTGTGGCGAAGAAAGAAATCCCG 394
134 sValSerGlySerAspThrMetSerPheGly..... 144
395 ATCAACATCGTTTACTTATAAATGTGAACGCAATATTATTAAGCA 444
145 .....TyrAlaLysAsnThrTyrThrAla 152
445 GGGACTAAGCCATCCTTATGCGGCGATTTATCATATCCCGCTTGCA 494
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
495 TAAATTTGCACAGATGCAGAAACCTGTTGAATGAC..... 531
169 rLysLeuValThrGluValAlaProAlaGluValSerAspIleGlyAlaV 186
532 .....AGTTATATGATGGGCGGAATATATGATCAAAATATATTAC 573
186 alSerGlyAlaTyrGlnAlaGly.....GlyArgPhe 196
574 CCTGACCGTGTCTATTGGCGCAGCAGCAATATGCGCGATGATGCA 623
197 ThrGluPheTyrArgLeuGlyGlyMetGlnTyrValLys..... 210
624 AGATGAGCCCATTAACCGGAAAGTTCATATCATATTGCAAGCGATAT 673
211 ..... 724
224 lPheLeuValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGly 240
724 .....ACAGTCACTTAGTATGTAAGAAAAATTAACATTAAGCCATA 764
241 GlnMetIleThrAlaGlnThrGly...AspIlePheLeuIleProProAs 256
765 TGGTTTTTACCAACAGAGGCGTCATTTGGCGAGCGGCTCCACCATGT 814
256 nGlyProLeuAlaAsnTyrIleuAsnMetGlyAspSerGlySerProLeuP 273
815 TTATGTATGATGCCAAAGCAAAAGTGTTAATTAATGGGATTTGCAA 864
273 heaIaTyrAspSerLeuGlnLysTyrValLeuIleGlyValLeuSer 289
865 ACGGCAACCCCTATATATGAAAAAGCAATGGCTCCAGCTGGTGTAA 914
.....|
290 SerGlyThrAsnTyr.....GlyAsnAsnTrpValValThrThrGln 303
915 AGATTGTTCTATGATGAATCTTGTGGAGATACCATCATGATTTCT 964
303 nAspPheLeu..... 306
965 ACGAACCACGTCAAATGGCAATACTTTTAACGACGATAATATGCG 1014
307 .....GlyGlnGlnProGlnAsnAsp..... 313
1015 ACAGAAAAATCATGCGCAACATGAACACATTTCTGCTATATGATT 1064
313 ..... 313
1065 AAAAAACAGAACCGTCAATGTTTATGTTCTTATTCGAGACAGCA 1114
314 .....PheAspLysThrIleAlaTyrThrSerG 323
1115 GAGAACCTGTT.....TATCATGCTGACAGTGGTGTCAAC... 1149
323 lYsGlnGlyValLeuGlnTyrLysTyrAspAlaAlaAsnGlyThrGlyThr 339
1150 .....AGTTATGCAACGACA..... 1164
340 LeuThrGlnGlyAsnThrThrTrpGlyTyrAlaTrpLysGluArgLysLe 356
1165 .....CTGATATATGAGAAAAATTTCTTATTTGCAAGAAAAAGCG 1210
356 uIleLeuAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyG 372
1211 AATTGATCTTACCAGCAACATCATCAAGTGTGCGAGATTAATTTTC 1260
372 lValValIleuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPhe 388
1261 CAAGAGATTTTACGGCTGCGCTGCAATTAACGAACCTGGCAAGGCG 1310
389 AlagIlyAspTyrAlaGlySerAlaLeuAsnGlyGlnThrTrpMetGlyG 405
1311 GGGCTTCATATCATGTAAGACAGTACCGTTACTTGAAGATTAACGGCG 1360
405 yGlyIleIleThrAspLysGlyThrHisValIleuTrpGlnValAsnGly 422
1361 TGGCAACGACCGCGTCCCAATCGGCAAGGCAAGCGCTGACGTTCAA 1410
422 alAlaGlyAspAsnLeuHisLysThrGlyGlnGlyLeuThrValAsn 438
1411 GCCAAAGGGAAAAACCAAGCTCGATCAAGCTGGCGCAGGTACAGTAT 1460
439 GlyThrGlyValAlaAsnAlaGlyGlyLeuLysValGlyAspGlyThrVal 455
1461 TTTGATCGACGAGGAGACGATTAAGGCAAAAAACAGCCTTTAGTAA 1510
455 eleuAsnGlnGlnAlaAspArgAspGlyLysValAlaIleAlaPheSerSer 472
1511 TGGCTTGGTCAAGCGGAGGAGTACGCTCACTGAATGCGCAATATCAG 1560
472 alGlyIleAlaSerGlyArgProThrValIleLeuSerAspSerGlnGln 488
1561 TTCACCCCGCAAACTATTTGGCTTGGCGCGGAGCGATTTGATTT 1610
488 ValAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyValLeuGlnIle 505
1611 AAACGGGATTCGCTTTCGTTCCACGCTATTCAAATAACGATGAAGGG 1660
505 uAsnGlyAsnAsnLeuThrThrPheThrArgLeuGlnAlaAlaAspTyrGly 522
1661 CGATGATTTGTACACACATCAAGACAAAGATTCACCGTTTACATT... 1707
522 lAlaIleIleThrAsn...AsnSerGlnLysLysSerThrValThrLeuAsp 537
1708 .....ACAGCATTAAGATATGCTAACACCGGCAATACAGAC 1748
538 LeuGlnThrLeuLysAlaSerAspIleAsnValProValAsnThrValSe 554

```

```
1749 CTTC.....GATACCAA. 1761
      |||
554 rtlepneglyargglyalaproglyaspleutyryaspsersert 571
      |||
1762 .....AAGAAATTGCTTACACGGTTGGTTGGC 1791
      |||
571 hrlysglnryrphelleuleylasersertyserpropheser 587
      |||
1792 GAG.....AAGATACGACCA 1808
      |||
588 Aspleuansanserseryaltrpqlansvalglylaspshiasnly 604
      |||
1809 AAC..... 1812
      |||
604 salalleasplhrVallyslnglnlyslleglualaserserglnpror 621
      |||
1813 .....AACGGCGCTCAACCTTGTTCACCAAGCC 1842
      |||
621 yrmetyrhlsglylnleuansnlyasmetaspvalasnlleprogl 637
      |||
1843 GCCGACGAGACCGCACCTGCTGCTTCGCGGCAACAATTA..AA 1889
      |||
638 leuserglyaspyalaleualaleuasplyserValasmlneurprogl 654
      |||
1890 CGGACATCAGCAACAACGCAACTGTTTTCAGCGGACAGCA 1939
      |||
654 uglyserlthrlyserglythrleuilepneglyrhlsprlv 671
      |||
1940 CACGCGACGCTTACATCATTTAAAGACCATTTGTCGCAAAAAGAGCG 1989
      |||
671 allehsalaglythrThrserserglnser..... 683
      |||
1990 ATTCCTCGCGGGAATCGTGTGGACACAGCTGATCAGCGCAT 2039
      |||
684 .....Asptrpelnthrarglnph 690
      |||
2040 TAAACGGAACCTTCAATTAAAGGCGACGCGGTGTTCCGCA 2089
      |||
690 ethrleuileuylsleuylsleuaspalalathrPhehlsleuserlrga 707
      |||
2090 ATGTCGCAAAAGTGAAGCGATTGGCATTTGACCAATCAGCCCAAGCA 2139
      |||
707 sr..glylseretnglnlyaspshiasnlatrhansglyserthrVal 722
      |||
2140 GTTTTGGTGTGCGACCCGATCAAAAGCCACACATCTGTACAGCTGGA 2189
      |||
723 ileuengly.....seretargvalPhehraspargserAs 735
      |||
2190 CTGGACGGGT...CTGACAAATGTGCAAAAAACCAATTACGACGATA 2236
      |||
735 pelythrnglyasnalaivalserSerValgluglyseralathrathrt 752
      |||
2237 AAGTATTCCTTCATTGACTAGACGACGACATCAGCGCATGTGCATCT 2286
      |||
752 hrVal.....glyaspGlnserAspTyrserglyasnValthrleu 765
      |||
2287 GCCGATCAGCTCATTTAAATCTCACAGGCTTGGCACATCAGCGCA 2336
      |||
766 gluaslylserSerleuInlle.....Metgluatgphethnglygl 780
      |||
2337 TCTTAGTCGCAAAATGGGATACGATTATACATCAGCCACAAGCCACC 2386
      |||
780 yllleglualatyraspsertThrValserValthrSerGlnasnalaValp 797
      |||
2387 AAAACGCAACCTTACCTGCGGCAATGCCCAAGCAACATTTAATCAA 2436
      |||
797 heasprargvalglyserPhevalasnSerSerleuthrleuenglyllysly 813
      |||
2437 GCCACATTTAAAGGCAACATCGGCTTGGCAATGCTTCATTTAACT 2486
      |||
814 Alalysleu.....ThralaglnserglyllepheserThrclal 827

2487 AACGACACAGCGCTTACAAAGCGAGTCTGACGCTTCCGCAACGCTA 2536
      |||
827 avalasp...VallysgluasnalaSerleuthrleuthrcllymethros 843
      |||
2537 AGGCAACGTAAGCCAT.....TCCGACCTCAACGCT 2568
      |||
843 erlathlnlysglnlytyrtyrserProvallleSerThrThrclugly 859
      |||
2569 AATGTCTCCCTAGCCGATAGGCACTATTCATTTGAACAGCGCGCTT 2618
      |||
860 ...ileasnleuclnlyspasnalaSerPne.....SerValylsas 872
      |||
2619 TACGCAACAATCAGCGCGGCAAGGATACGCAATTACACTTAAAGACA 2668
      |||
872 nmetglytyrleuserSer..... 878
      |||
2669 GCGAATGACGCTGCCGTCAGCAGCGAATTAGCAATTTAACTTGAC 2718
      |||
879 .....AsprlehsalaglythrThrAlalathrllleasnleuclly 892
      |||
2719 AAGGCCACATTCACCTCAATTCGCGCTATCGCACGATGCGGAGGCG 2768
      |||
893 Asp.....SerAspAl 896
      |||
2769 GCAAAACGCGAGTGCAGACAGATCGCCGCGCGCTTG...CGCCGCT 2815
      |||
896 aasplalaglylys...ThrAspserProleupheserSerleuMetlysg 912
      |||
2816 CGCGCCCTTCCCTAATTA...TCCGTTACACCGCACTTCGCTGAGAA 2859
      |||
912 lytyrAsnalaValleuargglySerllethrcllyalaglnserthrVal 928
      |||
2860 TCCGCTTCAACACGCTGACGCTAAC.....GCGAA 2891
      |||
929 AsnmetlleasnalaValleuThrPtyrserAspglyllysSerclnalacllyal 945
      |||
2892 ATTGACGCGTCAGGAGACATTCGCTTATGTCGCAACTTTCGCTTAC 2941
      |||
945 aleuylsAlalysglyserargllegluleuclnlyasprglyllyrhlsPhea 962
      |||
962 latrhleuclnlyVallygluLeuseralaspasnthrThrPheleuMet 978
      |||
2992 GCGGTCAACATACCAGGACAGCAACCTGCAAGCTGCAACATTTGACG 3041
      |||
979 HlsThrAsnasn.....SerargAlaspelnleuasnva 990
      |||
3042 AGTGGAA.....GGAAGACAAACAAACGCTGTCGAAAACCTTA 3082
      |||
990 lthrAsplylserglyserAsnanserValleuValAspPheleuA 1007
      |||
3083 ATTTCAACCTGCAAAAACGACACGTCGATGCGGCGCTGTACCA 3132
      |||
1007 sr..... 1007
      |||
3133 CTCATCCGCAAAAGCGGAGTTCGCGCTGATATCCGTCGCAAGACA 3182
      |||
1008 .....LysProAlaserglume 1013
      |||
3183 AGAGCTTCCGACAAACTCGGCAAGGCAAGGCCAAAAAACAGCGGAAA 3232
      |||
1013 tserValthr..... 1016
      |||
3233 AAGACACGCGCAAAACCTTGACGCGCTGATGCGGCGGCGGAGATCC 3282
      |||
1017 .....LeuilethrAlaProlyglylSer 1024
      |||
3283 GTCGAAAGACAAAGGCTTCCGACACGCGCGGCGGAGCGGAGGA 3332
      |||
1025 AspglulysThrPheThr.....AlaglyThrngl 1034
      |||
3333 AATGTGCGCATTTATGCAAGCGGAGAAAGAAAAAACGGGTGACAGCGG 3382
```



```

498 ATTGTCACAGATGCAAGACCTGTTGAAATACCACTATATGATGGG. 546
|||||..... |||..... |||
107 sPheValThrGluValAlaProIleLeaProThrAspAlaGlyGlyL 124
|||||..... |||..... |||
547 ..CGGAATATATCGATGCAAAATATACCGCGCGCTGTCGTTGGG 594
|||||..... |||..... |||
124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140
|||||..... |||..... |||
595 GCAGCGCAGCA.....TATTGCGCATGATGAAATGAGCCCAATA 638
|||||..... |||..... |||
141 AlagLysArgGlnLeuValTyrGlnLysGlyAlaTyrHisGlnGlyLys 157
|||||..... |||..... |||
639 CCGCGAAGTTCATAT.....CATATGCGAAGTGGTATCTGGC 679
|||||..... |||..... |||
157 n...GlnLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173
|||||..... |||..... |||
680 TCGTTGGTGGC.....AAT 693
|||||..... |||..... |||
173 lAlleAlaGlyThrProTyrLysAspLysAlaSerGlnIleAspGlnThrMetAsn 189
|||||..... |||..... |||
694 ACCTTTCACAAATGAGTATGATGATGCGCAGTCACTTATGATGTA 743
|||||..... |||..... |||
190 ThrGlnGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAla 206
|||||..... |||..... |||
744 AAAAATTAAACATAGC...CCATATGCTTTTACCAACAGGAGGCTCAT 790
|||||..... |||..... |||
206 uGlnLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223
|||||..... |||..... |||
791 TTGGGACATGCTGCACCAATGTTTATCTATATGCCCAAAAGCAAA 840
|||||..... |||..... |||
223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239
|||||..... |||..... |||
841 TGGTTAATTAATGGGATATGCAACGGCAACCCCATATATAGAAAG 890
|||||..... |||..... |||
240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys 254
|||||..... |||..... |||
891 CAATGCGTTCACAGCTGCTGTAAGATGCTTATGATGAATCTTG 940
|||||..... |||..... |||
255 .....LysSerTrpGlnGlnIleTrpAsnIleTyr. 263
|||||..... |||..... |||
941 CTGAGATACCCATTCAGTATTTACGAAACACGCTCAAAATGGGAATAC 990
|||||..... |||..... |||
264 .....LysLys 265
|||||..... |||..... |||
991 TCTTTTAAAGC.....GATATAATGGCACAGGAAAT 1025
|||||..... |||..... |||
266 GluPheAlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysG 282
|||||..... |||..... |||
1026 CAATGCCAAACATGACACAAAT.....TCTGCTCTAATGATTAA 1066
|||||..... |||..... |||
282 yAsnArgGlnHisIleTrpAsnIleThrPheGlyThrAsnSerLysIleG 299
|||||..... |||..... |||
1067 AAACAGCAACCGTCAATGTTTATGTTCTTTATCCGAGACAGCAAG 1116
|||||..... |||..... |||
299 LysThrAlaValArgLeuAlaAsn.....AsnGlnArg 310
|||||..... |||..... |||
1117 GAACCTGTTATCATGCTGACAGGTGTGTCACAGATTATGACCAGACT 1166
|||||..... |||..... |||
311 Asp.....Al 312
|||||..... |||..... |||
1167 GAATATGAGAAATATTTCTTATTATGACGAAGAAAGCGAATGTA 1216
|||||..... |||..... |||
312 AsnAsnGlnGlyAsnValThrPhe.....GlnAspAsnGlyThrLeuV 327
|||||..... |||..... |||
1217 TACTTACCGCAACATCATCAAGTGTGCGAGGATATATTATCCAAAGA 1266
|||||..... |||..... |||
327 AlLeuAspGlnAsnIleAsnGlnGlyAlaGlyLeuPheLysGly 343
|||||..... |||..... |||
1267 GATTTACGCTCGCTGAAATTAAGAA...ACTTGCACAAAGCGCGG 1313
|||||..... |||..... |||
344 AspTyrThrValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaG 360
|||||..... |||..... |||
1314 CGTTCAATCATGTAAGACAGTACCGTTACTTGGAAATTAACGGCGTG 1363
|||||..... |||..... |||

```

```

|||||..... |||..... |||
360 yLleAspValAlaAspGlyLysValLysValIleTrpGlnValLysAsnProA 377
|||||..... |||..... |||
1364 CAACAGCAGCCGCTGTCACAAATCGGCACAAAGCGACGCTGACGTTCAACC 1413
|||||..... |||..... |||
377 snGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGlnIleAsnGly 393
|||||..... |||..... |||
1414 AAGGGGAAACCAAGCTCGATGACAGCTGCGCGACGCTACAGTCATTTT 1463
|||||..... |||..... |||
394 ThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLe 410
|||||..... |||..... |||
1464 GGATTCAGCAGCAGACGATTAAGCAAAACACACCTTTAGTGAATCG 1513
|||||..... |||..... |||
410 uAsnGlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValG 427
|||||..... |||..... |||
1514 GCTTGGTCACGCGGAGGCTGACGCTGACGATGATGATGATGATGATG 1563
|||||..... |||..... |||
427 LysIleValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnIleLe 443
|||||..... |||..... |||
1564 AACCCGCAACACTTATTTGCGCTTTCGCGCGGACGTTTGATTTAA 1613
|||||..... |||..... |||
444 AsnProAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAs 460
|||||..... |||..... |||
1614 CGGGCATTCGCTTTCGTTCCACCGTATTCACCAATACCGATGGAAGGCGGA 1663
|||||..... |||..... |||
460 nGlyAsnAspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaA 477
|||||..... |||..... |||
1664 TGATTCACACCAACATTCACACCAAGAAATCCACGTTACCTTACAGGC 1713
|||||..... |||..... |||
477 rGlyLeuAlaAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGly 493
|||||..... |||..... |||
1714 AATAAGATATGCTACACACGCGCAATACACACGCTTG 1752
|||||..... |||..... |||
494 ...LysSerLeuIleThrAlaProGlnAsnLeuSerVal 505
|||||..... |||..... |||

```

seq_name: p1r2:S61315

seq_documentation_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) - *Neisseria meningitidis* (strain ENS3)

N:Alternate names: Iga proteinase; immunoglobulin A1 proteinase

C:Species: *Neisseria meningitidis*

A:Variety: strain ENS3771; strain ETH2

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S61315; S61316

R:Lomholt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neiss*

A:Reference number: S61314; MUID:95302961

A:Accession: S61315

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOM>

A:Cross-references: EMBL:X82468; NID:g732865; PIDN:CA57851.1; PID:g732866

A:Experimental source: strain ENS3771

A:Accession: S61316

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOM>

A:Cross-references: EMBL:X82469; NID:g732887; PIDN:CA57852.1; PID:g732888

A:Experimental source: strain ETH2

A:Gene: iga

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

alignment_scores: 886.00 Length: 541

Quality: 2.454 Gaps: 19

Percent Similarity: 66.728 Percent Identity: 39.187

alignment_block:

us-09-303-518d-649 x S61315

Align seg 1/1 to: S61315 from: 1 to: 496

```

202 GCGAAGATATGAGGTTTACAGAAAAAGGAGATGTCGGCAATC 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 AlatrAspLeuSerValLysAsnLysGlnGlnAsnIleGlnAsnI 17
252 AATGACAAAAGCCCCGATGATGATGATGATGATGATGATGATG 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 aleuSerAsnValProMetIleAspPheSerValAlaAspValAsnArg 34
299 GCGGCGGCGCATGTCGTGGCGCATATATATGTCGTCGTCGTCG 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 rGlnIleuThrValIleAspProGlnIleValIleValSerValLysHisVal 50
349 GCGGCTATATACAGCTGATGATGATGATGATGATGATGATGAT 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 LysGlyAspGlnIleSerIleGlnIleHisAsnGlnIleHisVal 67
397 ...CAACATCGTTTACTTATATAATGTCGAAACGCAATATATTAAG 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 lSerAsnAspLysAsnGlnIleIleArgSerValAlaGlnAsnAspTyrGln 84
443 CAGGACTAAAGGCATCCTTATGCGGCG...GATTAT 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 roAsnLysAsnIlePheHis...HisGlyAsnGlnIleArgLeuGlnAspTyr 99
478 CATATGCCGCTTTCATATAATGTCACAGATGACAGAACCTGTCGAAT 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
100 AsnMetAlaArgLeuAsnLysPheValThrGlnValAlaProIleAlaIle 116
528 GACCATATATGATGGG...CGCAATATATCGATCAATATTAATACC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 oThrSerAlaGlyGlyIleValIleThrTyrLysAspLysAsnArgPheS 133
575 CTACACGCTGTCGATGTCGCGAGCGACCAATATGCGCATGTATGA 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 eArgIleValAlaArgValIleGlyIleThrGlnPhe..... 144
625 GATGACCCCAATACCGCAAGTTCATATGATGTCGATGTCGATGTC 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ...GluTyrAsnSerArgTyrAsnMetThrGlnIleuSerArgAlaTyrAr 160
675 TTGGCTGCTGTCGCAATACCTT.....GCACAAATG 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
160 gTyrAlaIleIleAlaGlyThrProTyrGlnAspValAsnValIleHisVal 177
710 GATCAGATGTCGACAGTCAACTAGTAGTGAATAAATTAACATAGC 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177 euAsnGlnIleGlyLeuIleGlyPheGlyAspAsnSerLysHisIleSer 193
760 CCA.....TATGTTTTTACCAACAGAGG 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 ProGlnLysLeuLysGlnValLeuSerIleAsnAlaLeuThrAsnTyrAl 210
786 CTCATTTGGCAGACAGTGCACCAATGTTATGATGTCGCAAAAGC 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 aValIleuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGlu 227
836 AAAAGGTTTATTAATGAGGATGTCGAAACGCGCAACCCCTATATAGA 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 ysArgTyrValPheLeuGlyAlaTyrAspTyrTyrTyrAlaGlyTyr..... 241
886 AAAAGCAATGCTTCACAGTGTGTCGTAAGATG...TTTCATGATGA 932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 GlnLysAsnSerTyrGln.....GluTyrAsnIleTyrTyrLys 254
933 AATCTTCTGAGATACCATTCAGTATTCAGAACACGTCGCAAAATG 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 sGlnPheAlaAsp.....GluIleLysGlnArg. 263
983 GGAATACTCTTTAAGAGATATATGTCACAGAAACCAATCAATGCC 1032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

264 .....AspAsnAlaGlyThrIleLysGlyTyrGly 273
1033 AACATGACACAAATTCCTCCCTAATATGATTAACACAGAACCTTCA 1082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 GlnHisThrIlePheLysThr.....ThrIleThr..... 282
1083 ATTGTTAATGTTTCTTATACCGACAGACGACAGAACCTGTTATCATG 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 .....AsnSerHisIleGlySerThrIleValAlaGlyLeuAlaGlyAsnG 297
1133 CTGACAGGTCGTCACAGATTCAGACCCAGACTGATGATATGAGAAAT 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 LysArgGlyAla.....AsnAsnGlyGlnAsn 305
1183 ATTTCTTTATGACAGAGAAAGCGAATGATGATTCACAGCAATC 1232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 ValThrPhe.....GluAspAsnGlyThrLeuValLeuAspGlnAsnI 320
1233 CAATCAAGGTCGTGGAGATATATATTCACAGAGATTTTACGTCGTC 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337
1283 CTGAATAATACGAA...ACTTGGCAGGCGCGCGCTCATATCACTGA 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 LysAlaAsnAsnAspIleThrThrPheGlyAlaGlyIleAspValAlaAsp 353
1330 GACATGACCTTACTTGAAAGTAAAGCGCGTGGCAACAGCCGCTGTC 1379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 GlyLysLysValValThrGlnValLysAsnProAsnGlyAspArgLeuAl 370
1380 CAAATTCGCGCAAGGCGACGCTGACGCTTCACAGCAAAAGGCAACAG 1429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 aLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnG 387
1430 GCTCGATCAGCGCTGGCGCAGCGTACATGATTTGATCAGCAGCAGAC 1479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 LysIleuLysValGlyAspArgLysThrValIleLeuAsnGlnIleAlaAsp 403
1480 GATTAAGGCAAAAAACAACCTTATGTAATCGCTTGGTCAGCGCAG 1529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420
1530 GGGTACGTCGCACTGATCGATGATGATGATGATGATGATGATGAT 1579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 gGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 437
1580 ATTTGCGCTTTCGCGCGGACGCTTGGATTTAAACGGGCATTCGCTTCG 1629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 yThrPheLysPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCACCGGATTCGAAATACCGATGAGGCGGATGATGTCACACGCA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAs 470
1680 TCAACACAAAGATCCACCGTTACCATTCACAGCAATATATGATGCTA 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 nThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIle 486
1730 CAACCGCAATACACACGCTTG 1752
486 nThrAlaProGlnAsnLeuSerVal 493

```

seq_name: p1r2:S61328
seq_documentation_block:
Iga-specific metalloendopeptidase (BC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: NG44/76
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61328
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss

A:Reference number: S61328; MUID:95302961

A:Accession: S61328

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-497 <LOW>

A:Cross-references: EMBL:X82481

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 883.50 Length: 546
Ratio: 2.461 Gaps: 18
Percent Similarity: 65.751 Percent Identity: 38.462

alignment_block:

US-09-303-518D-649 x S61328 ..

Align seg 1/1 to: S61328 from: 1 to: 497

```

202 GCGAAGATATGTAGGTTTACACAAAAAGGAGTGTGCGCAATC 251
    ||| |||:||||| |||:|||||: |||:|||||: |||:|||||:
1 AlathraspleuserValyllysasnlysglnlyglnasnlylelyasnly 17
252 AATGACAAAAGCCCGATGATGATTTCTGCTGCTGCGGTAAC..G 298
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 aleuserasnlyalprometileasprheserValalaaspvalasnlyg 34
299 GCGGTGGGCGATGTGGGCGATCAATATTTGTGACGCTGGCACAATAC 348
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 rghlthleuthvaliliasprglnlyalvalserValyllyshnlyval 50
349 GCGCGCTATAACAAGCTGATTTTGTGCGCAAGAGAAATCCCAT.. 396
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
51 lysglysprglnlyserlytytyglnlyshnlyasnlyshnlyleuasp 67
397 ....CAACATCGTTTACTATTAATTTGCAACGGAATTAATTTAAAG 442
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 lserasnspglnasnlytyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 84
443 CAGGACATAAAGGCCATCCTTATGCGCGC.....GATTAT 477
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 roasnlyasnlytrpHis...Hislyasnlyglnlytyrtyrtyrtyrtyr 99
478 CATATGCCCGCTTGCATTAATTTGTGACAGATGACAGACCTGTTAAAT 527
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 asnmetalaargleuasnlysphevalthrlyglnvalalprolilealpr 116
528 GACCAATTATATGATGGG...CGAATAATATATGATCAAAATATATAC 574
    ||| ||| :|||:|||||:|||||:|||||:|||||:|||||
116 thrserlealaglyglnlyvalglnlythrlytyrtyrtyrtyrtyrtyr 133
575 CTGACCCGCTTGTATTTGGGCGCAGGCAATTTGGCGCATGATGAAGA 624
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 erglrphavalargvalylalaglythrlyglnphe..... 144
625 GATGAGCCCAATACCGCAAGGATTCATATGATATGCAAGTGGCATTC 674
    ||| ||| :|||:|||||:|||||:|||||:|||||:|||||
145 ...glutyrasnseralgtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 160
675 TTGGCTGCTGGTGGCAATACCTT.....GCACAAATG 709
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 gtyralalilealaglythrlythrlyglnasprvalasnvalthrserasn 177
710 GATCGAGTGTGGCGACAGTCACTAGGATGAGAAATTAACATATGAC 759
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
177 euasnlyglnlyleuileglyrphleglytyrtyrtyrtyrtyrtyrtyr 193
760 CCA.....TATGTTTTTACCAACAGGAGG 785
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
194 progllyleuileuysglnvalleuserglnasnalaletuThrAsnlyval 210
786 CTCATTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835

```

```

: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 avallleuglyaspserglyserproleuphealalyasprlysglnlyul 227
836 AAAAGTGTATTAATGAGGATGATTCGAACGGCAACCCCTATATATGGA 865
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
227 ysargtrpvalrphleuglyalatyrsprlytyrpalaglytyr..... 241
886 AAAAGCAATGCGCTCCACCTGCTGTAAGATGAG...TCTCTGATGGA 932
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
242 glnlyasnsertrpeln.....glutyrasnlylelytyrlygly 254
933 AATCTTGTCTGAGATACCATTCAGTATCTACAGACCAACGTAATG 982
    : ||| ||| :|||:|||||:|||||:|||||:|||||:|||||
254 sglurphealaspriyllely..... 261
983 GGAATATCTCTTTTAACGACGATATATGCGACAGAAATATCATGCGC 1032
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 .....glnaraspasnalaaglythrlylelysglyasnly 273
1033 AATCATGAACACAT.....TCTGCTGCTAATTAATTAACACG 1073
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 glnlyshnlytrpasnlylethrphleglythrlyasnserlylleglyserth 290
1074 AACCGTTCAATTTGTTAATGTTCTTTATCGACAGACAGACAGACCTG 1123
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 ralaValargleu..... 294
1124 TTTATCATGCTGCGAGTGGTGCAACAGTTATCGACCCAGACTGAATAT 1173
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
295 .....Alaglyasnlyglnlyasp.....Alasnlysn 303
1174 GAGAAATATATTTCTTTATTCGACGAGAAAGCGAATATATCTTAC 1223
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
304 glylinsnvalthrph.....gluaspasnlythrlyleuvalleuasn 318
1224 CAGCAATCATCAAGGCTGAGAGATTAATTTCCAGAGATTTTA 1273
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 pglinsnlyleasnlylelyalaglyleuphepelysglyaspptyr 335
1274 CGGTCTGCGCTGAATAATACGAA...ACTTGGCAAGCGCGCGCTTCAT 1320
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
335 hrvallysglyleasnlyasnlylethrtrpelyglnlyleaspr 351
1321 ATCAGTGAACAGATACCTTACTTGGAAATGAAAGGGGTGGCAACCA 1370
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 Valthrpsglnlytyrlysnlyvallythrlyglnlyleasnlytyrthrly 385
1371 CCGCTGCTCCAAATCGCAAGGCAAGCGTGCAGCTTCAAGCCAAAGGG 1420
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 pargleualalyrlylelyglnlythrlyglnlyleasnlytyrthrly 385
1421 AAAACCAAGCTGATACGCTGGGCGAGCGTACAGTATTTGGATCGAG 1470
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
385 alasnlyglnlyleuileuysvalyllyaspelythrlyvallyleuasnly 401
1471 CAGCGACAGATTAAGGCAAAACCAAGCCTTAACTGTAATTCGCTGGT 1520
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 glnalaaspvalaspriyllyvalglnalrphserglnvalilyleval 418
1521 CAGCGGACAGGGTACGCTGACATGATGCCGATATCACTTGAACCCCG 1570
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 lserglyarglythrlyleuvallyleuasnserasnlyshnlyleasnly 435
1571 ACAACCTATATTTGGCTTGGCGGAGCTTGTGATTAACGGGCAAT 1620
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
435 spasnlyleuylrphleglythrlyglnlytyrtyrtyrtyrtyrtyrtyr 451
1621 TCGCTTGTCTCCACCTATTCAAATACGATGACGAGGCGATGATGT 1670
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 asprlythrphleglnlyshnlyleuasnlyvalaspelyalalyaleval 468
1671 CAACCAATCAAGCAAAAGATTCACCGTTACCAATTAACAGCAATTAAG 1720
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```


468 laenhisasnhrSerHisAlaSerThrIleThrLeuThrGly...Lyss 484

1721 ATATGCTACACCGGCAATACACAGCTTGATAC 1758

464 erleuilethrAsnProAsnSerLeuSerValHisSer 496

seq_name: p1r2:S61323

seq_documentation_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) - *Neisseria meningitidis* (strain HP48

N:Alternate names: Iga protease; Immunoglobulin A1 protease

C:Species: *Neisseria meningitidis*

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

R:Accession: S61323, S61324

R:Domolt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseri*

A:Reference number: S61314; MUID:95502961

A:Accession: S61323

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOM>

A:Cross-references: EMBL:X82480; NID:9732875; PIDN:CA57863.1; PID:9732876

A:Experimental source: strain HP48

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOM>

A:Cross-references: EMBL:X82475; NID:9732879; PIDN:CA57858.1; PID:9732880

A:Experimental source: strain HP96

C:Genetics:

A:Gene: iga

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 878.00 Length: 543

Ratio: 2.419 Gaps: 19

Percent Similarity: 66.851 Percent Identity: 38.858

alignment_block:

US-09-303-518d-649 x S61323 ..

Align seg 1/1 to: S61323 from: 1 to: 496

202 GCGAAGATATGAGTTTACACAAAAGGAGTTGTGCGCAATC 251

1 AlatrhrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17

252 AATGCAAAAGCCCGATGATTTCTGTGTCGTGCGGTAC...G 298

17 AleuSerAsnValPrometIleAspPheSerValAlaAspValAsnArg 34

299 GCGTGGCGCATGTGGTGGCGCATATATATTTGAGCGTGCACATAAC 348

34 rGthLeuThrValIleAspProGlnThrValAlaSerValLysVal 50

349 GCGGCTATTAACAACGTTGATTTGGTGGCGAAGAGAAATCCGAT... 396

51 LysGlyAspGluIleSerTYTTrGlyHisHisAsnGlnHisLeuAspVa 67

397 ...CAACATGCTTTACTATTAATAATGTAAACGGAATATTAATAAG 442

67 lSerAsnAspGlnAsnGlnThrValArgSerValAlaGlnAsnAspTyrGlu 84

443 CAGGACTAAAGGCCATCCTTATGCGGC...GATTAT 477

84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlnThrValGluGlnAspTyr 99

478 CATATGCCGCGTTTCATAATTTGCACAGATGACGAACCTGTGAAT 527

100 AsnMetAlaArgLeuAsnLysPheValThrGlnValAlaProIleAlaPr 116

528 GACCAGTTATATGATGGG...CGGAATATATCATCAAAATATATAC 574

116 oThrSerAlaGlyGlyValGlnThrTrpLysAspLysAsnArgPhe 133

575 CTGACCGTGTCTGATTTGGGCGAGCGAGCAATATTTGGCATGTGTA 624

133 erGluPheValArgValGlyAlaGlyThrGlnPhe..... 144

625 GATGAGCCATATACCCGGAATGTCATATCATATTTGCAATGCTATTC 674

145 ...GluTrpAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyr 160

675 TTGGCTGTGGTGGCAATACCTTT.....GCACAAATG 709

160 gTyAlaIleAlaGlyThrProTrpGlnAspValAsnValThrSerAsn 177

710 GATGAGTGTGCGACAGTCACCTTAGTATGTAATAAATTAACATAGC 759

177 euAsnGlnGlnGlyLeuIleGlyPheGlyAspAsnSerLysHisSer 193

760 CCA.....TATGGTTTTTCCACACAGAG 785

194 ProGlnLysLeuLysGlnValLeuSerGlnAsnAlaLeuThrAsnTrpAl 210

786 CTCATTTGGCGACAGTGGCTCACCAATGTTATCTATGATGCGCAAAAGC 835

210 ...GluLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGlu 227

836 AAAGTGTATTAATTAATGGGTATGGCAACGGCAACCCCATATATAGA 885

227 ysaArgTrpAlaPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241

886 AAAGCAATGCGCTCCAGCTGGTGTGAAGATTGG...TTCTATGATGA 932

242 GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLysLys 254

933 AATCTTGTGGAGATACCATTCAGTATTTACAGCACCGCTCAAAATG 982

254 sGluPheAlaAsp.....GluIleLysGlnArg. 263

983 GGAATACTCTTTTAACGACGATTAATGACACAGCAAGAAATCAATGCC 1032

264AspAsnAlaGlyThrIleLysGlyTyrGly 273

1033 AAACATGAAACAAATTTCTGCTTAATGATTAATAAACACAGCAACCTTCA 1082

274 GluHisHisTrpLysThr.....ThrGlyThr..... 282

1083 ATTGTTAATGTTCTTATCCGAGACAGCAAGAACCTTTATCATG 1132

283AsnSerHisIleGlySerThrAlaValArgLeuAlaGlyAsnG 297

1133 CTGACGGTGGTGTCAACAGTTATGACCCAGACGATTAATGAGAAAT 1182

297 lUaTrGlyAla.....AsnAsnGlnGlnAsn 305

1183 ATTTCCTTATTAAGCAAGAAAGGCAATGATTAATTTACCAACAT 1232

306 ValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsn 320

1233 CAATCAAGTGTGCTGAGATTTATTTCAAGAGATTTTACGCTGCC 1282

320 eaSnglnGlyAlaGlyLeuPhePheLysGlyAspTyrThrValLysG 337

1283 CTGAAATTAACGAA...ACTTGGCAAGCGCGCGCTCATATCACTGAA 1329

337 lYlleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

1330 GACAGTACCTTACTTGAAGAGTAACAGCGGTGCAACAGCGCGCTGC 1379

354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370


```

1183 ATTTCTTTATTGACGAGGAGGAGGATTTGATCTACCAACAT 1232
      :::::||||| :::::||||| :::::||||| :::::|||||
306 ValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsn 320
1233 CAATCAGAGGCTGAGAGATTTATTTCCAGAGATTTTACGGTCTGC 1282
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
320 easnGlnGlyValGlyLeuPhePheGlyAspGlyThrValValysG 337
1283 CTGAAATATAGAA...ACTGGCAGGGGGCGCTTCATATCACTGAA 1329
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
337 LylLeuAsnAsnAspIleThrTrpLeuGlyAlaGlyLeuAspValAlaAsp 353
1330 GACATACGCTTACTTGGAAAGTAAACGGCTGGCAACGACGGCTGTC 1379
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
354 GlyLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370
1380 CAATATCGGAAAGCAGCGCTGCAGCTTCAACCCAAAGGGGAAACCAAG 1429
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
370 AluYsIleGlyLysGlyThrLeuGlnLeuAsnGlyThrGlyValAsnGlnG 387
1430 GCTCGATCAGCGTGGCGAGCGTACAGTCATTTGGATCAGCAGCGCAGC 1479
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
387 LylGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAsp 403
1480 GATAAAGCAAAAACACACCTTGTAGTGAATCGCTTGTGACGCGCAG 1529
      :::::||||| :::::||||| :::::||||| :::::|||||
404 SerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420
1530 GGGTACGGTCAACTGAATGGCGATATACAGTTCACCCGCAAACTCT 1579
      |::|::|::|::|::|::|::|::|::|::|::|::|::|
420 garGhrIleuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuT 437
1580 ATTTGGCTTTTCGCGCGACGCTTGAATTAACGGCATTCGCTTTCG 1629
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
437 yrrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCCACCCGATTCAAATACGATGAGGCGCGATGATGTGCAACCAAA 1679
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAs 470
1680 TCAGACAAAGAAATCACCGCTGATATACAGCAATTAAGATATGTCTA 1729
      |::|::|::|::|::|::|::|::|::|::|::|::|::|
470 nThrGlyHisThrSerThrIleThrLeuThrGly...LysSerLeuIleT 486
1730 CAACCGCATATACACAGCTTGATAGC 1758
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
486 HisAsnProAsnSerLeuSerValHisSer 495
seq_name: p1z:S61329
seq_documentation_block:
  Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
  C:Species: Neisseria meningitidis
  A:Variety: NGC16
  C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
  C:Accession: S61329
  R:Jomholt, H.; Poulsen, K.; Mogens, K.
  Mol. Microbiol. 15, 495-506, 1995
  A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseri
  A:Reference number: S61314; MUID:95302961
  A:Accession: S61329
  A:Status: preliminary; nucleic acid sequence not shown
  A:Molecule type: DNA
  A:Residues: 1-508 <10M>
  A:Cross-references: EMBL:X82485; NID:g732869; PIDN:CAAS7868.1; PID:g732870
  C:Superfamily: Iga-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
  Quality: 870.00      Length: 553
  Ratio: 2.465        Gaps: 22

```

Percent Similarity: 63.834 Percent Identity: 39.060

alignment_block:

US-09-303-518D-649 x S61329

Align seg 1/1 to: S61329 from: 1 to: 508

```

202 GCGAAGATATTTAGGTTTACCAAAAAAGGAGATTGGTCGCAATC 251
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAAAGCCCGATGATGATTTTCTGTGGTGGCTAAC...G 298
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
17 AluSerAsnValProMetIleAspPheSerValAlaAspValAsnLysSA 34
299 GCGTGGCGCATGTTGGTGGCGATATATATGTGTGAGGTGGCAT... 345
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
34 rGlyLeuAlaThrValValAlaAspProGlnTrpAlaValSerValLysHisAla 50
346 .....AACGGCGCTATTAACAACGTTGATTT 371
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
51 LysAlaGlnValHisThrPheThrGlyGlnGlyGlnLysAsnAsn 67
372 TGGTGGCGAAGAGAAATCCCGATCAACATCGTTTCTTATTAATTG 421
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
67 rValAlaAspLysGlyAsn.....GluTrpArgValY 78
422 TGAAACGATATATTAATAAGCAGGAGCTAAAGCCATCTTATGAGGGC 471
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
78 AlGlnGlnAsnAsnTrpGluPro.....HisLysAlaTrpGly 90
472 .....GATTATCATATGCCGCTTTGCATAA 497
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
91 AlaSerAsnLeuGlyArgLeuGlnLysPheTrpGlnMetAlaArgPheAsn 107
498 ATTTGTCACAGATCGCAGACCTGTTGAATAGCAGATTTATGATGGG. 546
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
107 sPheValThrGlnValAlaProIleAlaProThrAspAlaGlyGlyL 124
547 ..CGGAATATATGATCAATAATTAATTAACCTGACCGTGTTCGATTTGG 594
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
124 euAspThrTrpLysAspLysAsnArgPheSerPheValArgValGly 140
595 GCAGCAGGCAAA.....TATTGGCGATCGATGAAATGACCCCAATA 638
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
141 AlaGlyArgGlnLeuValTrpGlnLysGlyAlaTrpHisGlnGlnGlyAs 157
639 CCGCGAAGATTCATAT.....CATATTGCAAGTGGCTATTCTTGGC 679
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
157 n...GluLysGlyTrpLysPheLeuArgAspLeuSerGlnAlaTrpArgTrp 173
680 TCGTTGGTGGC.....AAT 693
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
173 lAlaIleAlaGlyThrProLysAspIleAsnIleAspGlnThrMetAsn 189
694 ACCTTTGACCAAAATGATGATGAGTGGTGGCAGTCAACTTACGTAGTGA 743
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
190 ThrGlnGlyLeuIleGlyPheGlyAsnHisAsnThrHisThrGlySerAla 206
744 AAAAATTAACATAGC...CCATATGCTTTTACCACAGAGAGCTCAT 790
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
206 uGlnLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTrpGlyVal 223
791 TTGGCGAGTGGCTGACCAATGTTATATCATATGATGCCAAAAGCAAAAG 840
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239
841 TGCTTAATTAATGGGATTTGCAACGGCAACCCCTATATGAGAAAG 890
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|
240 TrpValPheLeuGlnThrTrpLysAspTrpTrpAlaGlyTrp...GlyLys 254
891 CAATGCTTCACAGTGTCTGTAAGATGCTTCTATGATGAATCTTTG 940
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

255 .....LysSerTrpGlnGluTrpAsnIleTyr. 263
941 CTGGAGATACCCATTCAGTATTCTACGACACGTCAGAAATGGAATAC 990
264 .....LysLys 265
991 TCTTTTACGAC.....GATTAATATGACAGGAAAAAT 1025
266 GluPheAlaLysIleLysGlnArgAspAsnIleGlyThrIleLysGln 282
1026 CAATGCCAATCAAGACACATTCCTGCTATATAGATAAAAACAGAA 1075
282 yaannglylnhishistrasn.....IleThrPheGlyT 294
1076 CCGTCAATGTTTAAATGTTTCTTATCCGAGACACAGCAAGAACCTGTT 1125
294 hr.....AsnSerHisIleGlySerThrAla..... 302
1126 TATCATCTCGAGTGGTGTCAACAGTTATCGACCCAGCTGAATATG 1175
303 ValArgLeuAlaGlyAsnGluArgAsp.....AlaAsnAsnGln 315
1176 AGAAAAATTTCTTATTTATGACGAGAAAGCGAATTGATCTTACCA 1225
315 yGlnAsnValThrPhe.....GluAspAsnGlyThrLeuValLeuAspG 330
1226 GCAACATCAATCAAGTGTGTGAGAGATTATTTCCAGAGATTTTACG 1275
330 lnaSnIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThr 346
1276 GTCTGGCGTGAATAATACGAA...ACTTGGCAAGCGCGGCGTCTTAT 1322
347 ValIysGlyAlaAsnAsnAspIleThrTyrLeuGlyAlaGlyIleAspVa 363
1323 CAGTGAACACAGTACCGTTACTTGAAGAAAGTAACGCGTGGCAACGACC 1372
363 lAlaAspGlyLysLysValValTyrGlnValLysAsnProAsnLysAsp 380
1373 GCGTGTCCAAATGGGCAAGGCGCGTGCACGTTCAAGCAAGGGGAA 1422
380 rGleuAlaLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyVal 396
1423 AACCAAGCGTATACGCGTGGCGAGCGGTACAGTCAATTTTGATCAGCA 1472
397 AsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLys 413
1473 GGCAGACGATAAAGCAAAAAACAAGCCTTATGAAATCGGCTTGCTCA 1522
413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430
1523 GCGGAGAGGGTACGCGTGCACACTGGAATGCCGATTCAGTTCAACCCGAC 1572
430 eGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 446
1573 AAACCTATTTTCGGCTTTCGGCGCGAGCGTGTGATTTAAAGGGCATTC 1622
447 AsnLeuTyrPheGlyPheArgGlyLysArgLeuAspAlaAsnGlyAsnAs 463
1623 GCTTTCGTTCCACCGTATTCAAATACGATGAGAGGGCGGATGATGCA 1672
463 pLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValA 480
1673 ACCCAATCAAGCAAGAAATCCACGTTACATACGACGCAATTAATAT 1722
480 snHisAsnThrGlyHisIleAsnThrIleThrLeuThrGly...LysSer 495
1723 ATTGCTACA 1731
496 LeuIleThr 498
seq_name: p1r2:S61327
seq_documentation_block:

```

```

Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: NCG80
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61327
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the Iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314, MIMID:95302961
A:Accession: S61327
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82479; NID:g732883; PIDD:CAA57862.1; PID:g732884
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

alignment_scores:
Quality: 865.00 Length: 534
Ratio: 2.437 Gaps: 20
Percent Similarity: 66.479 Percent Identity: 39.326

alignment_block:
us-09-303-518d-649 x S61327 ..
Align seg 1/1 to: S61327 from: 1 to: 496

202 GCGAAGATATGAGGTTTACACAAAGGAGGAGTTGTGCGCAATC 251
1 AlarhAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17
252 AATGACAAAGCCCGATGATGATTTTCTGTGCTGCGGTAAAC...G 298
17 alauSerAsnValPrometIleAspPheSerValAlaAspValAlaAsnAlaG 34
299 GCGTGGCGGATTTGGTGGCGGATCAATATATGAGGTGCGGACATAC 348
34 rGthrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50
349 GCGGCGTATACACAGCTGATTTTGTGCGGAGAGAAATCCGAT... 396
51 LysGlyAspGlnIleSerTyrGlyHisHisAsnGlyHisLeuAspVa 67
397 ....CAACATCGTTTACTTATTAATTTGTGAACGGAATTAATTAAG 442
67 lserAsnAspGlnAsnGlnTyrArgSerValAlaGlnAsnAspTyrGln 84
443 CAGGACATAAGGCGCATCTTATGGCGGC.....GATTAT 477
84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGlnAspTyr 99
478 CATATGCCGCGTTTGATTAATTTGTACAGATGACGAACCTGTGAAT 527
100 AsnMetAlaArgLeuAsnLysPheValThrGlnValAlaProIleAlaPr 116
528 GACCAGTTATATGATGG...CGGAATATATCATGATCAAAATTAATAC 574
116 oThrSerAlaGlyGlyValGlnValGlyThrGlnPhe..... 144
575 CTGACCGTGTGATTTGGGCGAGCGGACGATATTTGGCATGTGAA 624
133 eGlnPheValArgValGlyAlaGlyThrGlnPhe..... 144
625 GATGAGCCCAATACCGCGAAGTTCATATCATATTTGCAAGTCGATTC 674
145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160
675 TTGGCTGTTGGTGGCAATACCTTT.....GCACAAATG 709
160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177
710 GATCAGGTGTGGCAGTCAGTACCTTAGTGTGAAAAAATTAACATAGC 759

```

```

437 ypehGgLyphearGgLygLyarGleuKspsrlaSnclYasInfspleuthr 453
1630 TTCACCCGCTATTCAAAATACCGATGAAGGGCGCATGTTGTCAACCACA 1679
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 pReGluHnIsIleargAsnValaSprGlyAlaIargIleValaShnIsAs 470
1680 TCAAGACAAGATCCACCGCTTACCATTTACAGCAATTAAGATATTGCTA 1729
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 mThrgIyHIsaIaserThrlleThrlleuthrgly...LysSerLeuIlet 486
1730 CA 1731
||
486 hr 486
seq_name: pIr2:S61320

```

```
alignment_scores:
  quality: 865.00
  ratio: 2.403
  percent_similarity: 64.516
  percent_identity: 38.889
```

Align seg 1/1 to: \$61320 from: 1 to: 508

```

202 GCGAAGGATATTGAGCTTTCCACACAAAAGGGGAGTTGGCCGCAATC 251
1 AlaIhrspIeuSerValIlysAsnIysGInIyGlnAsnIIleGIyAsnAl 17
252 AATGACAAAAGCCCCGATGTGATTTTCTGTGGTTCGGGTAAAC...G 298
17 aleuSerAsnValIprIoxetIIleSpheSerValAlaIAspAlaIAsnIySa 34
299 GCGGCGGGCGCATTTGGTGGGCGCATATATATTGACGGTGGCACA... 345
34 rGIleAlaIhrValIAspProGInIyAlaIValSerValIyGInIAla 50
346 .....AACGGCGGCTATACCAACGTTGATTT 371
51 LysAlaIleValIhIsthrPheIyTrIyGInIyAsnGInIyAsnAs 67
372 TGGTGGCGGAAGAGAATCCGCGATCAACATCGTTTACTATTAATTTG 421
67 pValAlaIAspIyGInIAsn.....GInIyAlaIyVal 78
422 TGAACGGAATATTATTAA.....GCAGGACTAAA 453
78 alGIuGInAsnAsnIyGIuIPronIhIysAlaItrPSerAlaIAsnIeu 94
454 GGCCATCTTATGGCGGCATATATATATGCGCGGTGGATTAATTTGT 503
95 GIyAlaIeu.....GInAspIyAsnMetAlaItrPheAsnIySpheVal 109

```

```

504 CACAGATGAGAACCGTTGAATGACACGATTATATGATGGG...CGGA 550
109 LTHRGIVAVIAIAPROIleAIPROThraspAlaGlyGlyLeuAspT 126
551 AATATTCGATCAAAATAATTAACCTGACCGGTTTCGATTGGGCGAGC 600
126 hTyrIlysaspIysasnArgPheSerSerPheValArgValGlyAlaGly 142
601 AGCGAA...TATTGGCGATCTGATGAAGATGACCCCAATAACCGGA 644
143 ArgGlnLeuValIyGlnIySglAlaTyrHISglnGlnIyAsn...GI 158
645 AAGTCATAT...CATATGCAAGTCGATTCCTGCTGCTGCTG 685
158 ulysgIlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleA 175
686 GTGGC...AATACCTTT 699
175 laGlyThrProTyrIlysAspIleAsnIleAspGlnThrMetAsnThrGlu 191
700 GCACAAAATGATCAGGTGGCAGACGATCACTTAGTAGTGAATAAT 749
192 GlyLeuIleGlyPheGlyAsnHISasnthrHisTyrSerAlaGlnIle 208
750 TAAACATAGC...CCATATGTTTTTACCAACAGAGGCTCATTTGGCG 796
208 ulysgIlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyA 225
797 ACAGTGGCTCAGCAATGTTATCTATGATGCCCAAAAGCAAAAGTGTTA 846
225 spSerGlySerProLeuPheAlaPheAspIysGlnIySasngIntPval 241
847 ATTAATGGGGTATTGCAACGGCACCCCTATATAGGAAAGCAATGG 896
242 PheLeuGlyThrTyrAspTyrTyrPAlaGlyTyr...GlyIys... 254
897 CTTCACACTGGTTCGTAAGATGGTCTATGATGAATCTTGGCTGAG 946
255 .....LysSerTyrGlnIuTyrPAsnIleTyr..... 263
947 ATACCATTCAGTATCTACGACACGTCGCAAAATGGAAATACCTTTT 996
264 .....LysIysGlnIle 267
997 AACGAC.....GATATTAATGGCCACAGAAATATCATGC 1031
268 AlaAspIySleIySglnArgAspAsnAlaGlyThrIleIySglnAsnGI 284
1032 CAACATGACACAACTCTCTGCTAATAGATTAAACACAGACCGTTC 1081
284 yGlnIHisTyrPAsn.....IleThrPheGlyThr... 294
1082 AATTGTTAATGTTCTTTATCCGACAGACAGACAGAACCTGTTATCAT 1131
295 .....AsnSerHisIleGlySerThrAla.....ValArg 304
1132 GCTGCAGGTGGTGCACAGTTATCGACCCGACTGATATATGGAGAAA 1181
305 LeuAlaGlyAsnGlnIyAsp.....AlaAsnAsnGlnIyGlnAs 317
1182 TATTCCTTTATGACGAAAGAAAGGCAATGTACTTACCGACAACA 1231
317 nValThrPhe.....GluAspAsnGlyThrLeuValIleuAspGlnAsnI 332
1232 TCAATCAAGTGTGAGGATTATATTTCACAGAGATTTTTACGCTCG 1281
332 leAsnGlnIyAlaGlyIyLeuPhePheIySglnIyAspTyrThrValIys 348
1282 CCGTGAATAACGAA...ACTTGGCAAGCGCGCGCTTCATATACGGA 1328
349 GlyAlaAsnAsnAspIleThrTyrPheGlnIyAlaGlyIleAspValAlaAs 365
1329 AACACAGTACCGTTACTTGAAAGTAAACGGCGTGCACAAACGACCGCTGT 1378

```

```

: ... |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 pGlyIyLysValValTyrpGlnValIyLysAsnProAsnGlyAspArgLeuA 382
1379 CCAAAATCGCGCAAGACGCGTCGACGCTTCACACCAAGGGAACCA 1428
382 laIySleGlyIyGlyThrLeuGlnIleAsnGlyThrGlyValAsnGln 398
1429 GCGTCGATCAGCGTGGCGGCGGATGATCATTTTGGATGACGAGCA 1478
399 GlyGlnLeuIySValGlyAspGlyThrValIleuAsnGlnIySAlaAs 415
1479 CGATTAAGGCAAAAACAGCGCTTTAGTGAATGCGCTGGTGCAGCGCA 1528
415 pSerAsnIyLysValGlnAlaPheSerGlnValGlyIleValSerIyA 432
1529 GGGTACGCTGCAACTGATATGCCGATATATCATTCACCCGACAACTC 1578
432 rGArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 448
1579 TATTTGGCTTTTCGCGGCGGACGTTTGGATTAAACGGCATTCGCTTC 1628
449 TyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuTh 465
1629 GTTCACCGCTATTCAAATACCGATGAGAGGGCGGTGTTGTCACACAA 1678
465 rPheGlnHisIleArgAsnValAspGlnGlyAlaArgIyIleValSnhISA 482
1679 ATCAAGACAAAGAAATCCACCGTTACCATTTACAGCAATTAAGATTCCT 1728
482 snThrGlyHisAlaSerThrIleThrLeuThrGlnIy...LysSerLeuIle 497
1729 ACNACCGCATATACACACGCTTG 1752
498 ThrAlaProGlnAsnLeuSerVal 505

seq.name: pir2.S61322

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: HF54
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61322
R:Lombolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61322
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOW>
A:Cross-references: EMBL:X82473; NID:9732877; PIDN:CAAS7856.1; PID:9732878
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase, metalloproteinase

alignment_scores:
Quality: 864.00 Length: 541
Ratio: 2.407 Gaps: 19
Percent Similarity: 66.359 Percent Identity: 38.632

alignment_block:
US-09-303-518d-649 x S61322 ..
Align seg 1/1 to: S61322 from: 1 to: 496

202 GCGAAGATTTGACGTTTACACAAAGGAGGAGTGGTGGCAATC 251
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AlatrAspLeuSerValIySAsnIySglnGlyGlnAsnIleIySAsnAl 17
252 AATGACAAAGCCCGGATGATTTTCTGCTGCTGCGCTAAC...G 298
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 aluSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

```

1133 CTCGACGGGTGTCAACAGTTATCGACCOAGACCTCAATATACGCAAAAT 1182
||||:|||||
297 lUArglyAla.....AsnAsnGlyGlnAsn 305
1183 ATTTCCTTATTGACGAAGAAAGCGCAATTGATCTACTACGACAAAT 1232
:||||:|||||
306 ValThrpe.....GluAsnAsnGlyThrValLeuAspGlnAsn 320
1233 CAATCAAGTGTGGAGATTATATTCACAGGAGATTTTACGGCTCGC 1282
|||||
320 eAsnGlnGlyAlaGlyGlyLeuPhePheGlyAspThrThrValLysG 337
1283 CTGAATAATACGAA...ACTTGCGAAGCGCGGGCGGTTCATCATGCTAA 1329
|||||
337 lYAlaAsnAsnGlyLeuThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353
1330 GACAGTACCGCTTACTTGGAAGTAACGGCGCTGGCAACGACCGCTGTC 1379
:||||:|||||
354 GlyLysLysValValThrGlnValLysAsnProAsnGlyAspArgLeuAl 370
1380 CAAAATCGGCAAGGACAGCTGACGTTCAAGCCAAAGGGAACCAAG 1429
:|||||
370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGln 387
1430 GCTGATCAGCGTGGCGGACGCTACAGTCAATTTGGATCAGCAGCGAC 1479
||||:|||||
387 lYglnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403
1480 GATTAAGGCAAAAAACAAGCCTTAGTGAATTCGCTTGGCCAGCGGAC 1529
|||
404 AlaAspLysLysValGlnAlaPheSerGlnValIleValSerGlyArg 420
1530 GGGTACGCTGCACTGAATGCCGTAATCAAGTTCACCCGACAACTCT 1579
|||
420 gArgThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeu 437
1580 ATTTCGGCTTTCGGCGGAGCGTTGGATTAAAGGGCGCAATTCGCTTCG 1629
|||||
437 yTrpGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCACCGCATTCAAAATACCGATGAAGGGCGCATGATTCACACCAG 1679
||||:|||||
454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470
1680 TCAAGCAAGAAATCCACGCTTACCATTAACGCAATAAAGATTTGCTA 1729
|
470 nTrpAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486
1730 CAACCGGCAATACACACGCTTG 1752
|||:|||||
486 hTrAlaProGlnAsnLeuSerVal 493
seq_name: pLR2:S61330
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: SM1166
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61330
R:Holmolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61330
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1508 <LOW>
A:Cross-references: EMBL:X82486; NID:9732860; PIDN:CAA57869.1; PID:9732861
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores: Length: 551
 Quality: 856.00
 Ratio: 2.404
 Percent Similarity: 64.610 Percent Identity: 38.657

alignment_block:
 US-09-303-518D-649 x S61330

Align seg 1/1 to: S61330 from: 1 to: 508

```

202 GCGAAATATTTGAGTTTACAAACAAAAGGGAGTTGGTCGCAATC 251
    ||| |||.....||| |||.....|||.....: |||.....:
1  AlathrSpleuSerValLysAsnLysGlnGlnAsnIleGlysnAl 17
    :|||.....: |||.....: |||.....: |||.....: |||
17 aleuSerAsnValPrometIleAspSerValAlaAspValAsnLysA 34
    :|||.....: |||.....: |||.....: |||.....: |||
252 AATGACAAAAGCCCGATGATTTCTTCTGCTGCGCTAAC...G 298
    :|||.....: |||.....: |||.....: |||.....: |||
17 aleuSerAsnValPrometIleAspSerValAlaAspValAsnLysA 34
    :|||.....: |||.....: |||.....: |||.....: |||
299 GCGTGGCGCATTTGGTGGCGATCATATATTTGGACGCGCACAT... 345
    :|||.....: |||.....: |||.....: |||.....: |||
34 rglIleAlaThrValValAspProGlnValSerValLysShsAla 50
    :|||.....: |||.....: |||.....: |||.....: |||
346 .....ACGGCGCTATACAAAGCTGATTT 371
    ||| |||.....: ||| |||.....: ||| |||.....: |||
51 LysAlaGlnValHisThrPheTyrGlyGlnTyrAsnGlnHisAsnAs 67
    :|||.....: |||.....: |||.....: |||.....: |||
372 TGTGCGGAAGAGAAATCCCGATCAACATCGTTTACTATTAATTTG 421
    :|||.....: |||.....: |||.....: |||.....: |||
67 pValAlaAspLysGlnAsn.....GlnTyrTrpValV 78
    :|||.....: |||.....: |||.....: |||.....: |||
422 TGAACGGAATATATATAA.....GACGAGACATAA 453
    :|||.....: |||.....: |||.....: |||.....: |||
78 alGlnGlnAsnAsnTyrGlnProHisLysAlaTrpSerAlaSerAsnLeu 94
    :|||.....: |||.....: |||.....: |||.....: |||
454 GGCATCTTATWGGCGGATTCATATGCGCGTTGCGATAAATTTGT 503
    :|||.....: |||.....: |||.....: |||.....: |||
95 GlnTrpGlu.....GlnAspTyrAsnMetAlaTrpPheAsnLysPheVa 109
    :|||.....: |||.....: |||.....: |||.....: |||
504 CACAGATCCAGACCTGTTGAATGACCACTTATATGATGGG...CGGA 550
    :|||.....: |||.....: |||.....: |||.....: |||
109 lThrGlnValAlaProIleAlaProThrAspAlaGlyGlyLysLeuAspT 126
    :|||.....: |||.....: |||.....: |||.....: |||
551 AATATGATCATCAAAATATATACCTGACCTGTTGATTTGGGCGGCG 600
    :|||.....: |||.....: |||.....: |||.....: |||
126 hTyrLysAspLysAsnArpPheSerSerPheValArgValGlyAlaGly 142
    :|||.....: |||.....: |||.....: |||.....: |||
601 AGGCAA.....TATTGGCATCTGATGAGATGAGCCCAATTAACCGCA 644
    :|||.....: |||.....: |||.....: |||.....: |||
143 ArgGlnLeuValTyrGlnLysGlyAlaTyrHisGlnGlnLysAsn...G 158
    :|||.....: |||.....: |||.....: |||.....: |||
645 AAGTTCATAT.....CATATGCAAGTGGCTATCTTGGGCGCTTG 685
    :|||.....: |||.....: |||.....: |||.....: |||
158 lLysGlnTyrAspLeuArgAspLeuSerGlnAlaTyrTrpTyrAlaIleA 175
    :|||.....: |||.....: |||.....: |||.....: |||
686 GTGGC.....AATACCTTT 699
    :|||.....: |||.....: |||.....: |||.....: |||
175 lalGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThrGln 191
    :|||.....: |||.....: |||.....: |||.....: |||
700 GCACAAATATGATCAGCTGGTGGCACGATCACTAGTAGGAGAAAAT 749
    :|||.....: |||.....: |||.....: |||.....: |||
192 GlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGlnGln 208
    :|||.....: |||.....: |||.....: |||.....: |||
750 TTAACATATAC...CCATATGTTTATTACCAACAGAGGCTCATTTGGCG 796
    :|||.....: |||.....: |||.....: |||.....: |||
208 ulysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValIleGlyA 225
    :|||.....: |||.....: |||.....: |||.....: |||
797 ACAGTGGCTCACCAATGTTATCTATGATGCCAAAGCAAAAGTGTTA 846
    :|||.....: |||.....: |||.....: |||.....: |||
225 spSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrpVal 241
    :|||.....: |||.....: |||.....: |||.....: |||
847 ATTATGGGCTATGCAAAAGGCAACCCCTATATAGAGAAAAGCAAGG 896
    :|||.....: |||.....: |||.....: |||.....: |||
242 PheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys..... 254
  
```

```

897 CTTCAGCTGGTTCGTAAGATGGTTCTATGATGAATCTTTCCTGAG 946
    :|||.....: |||.....: |||.....: |||.....: |||
255 .....LysSerTrpGlnGlnTrpAsnIleTyr..... 263
    :|||.....: |||.....: |||.....: |||.....: |||
947 ATACCATTCAGTATTCACAGACACGCTCAAAATGGCAATATCTTT 996
    :|||.....: |||.....: |||.....: |||.....: |||
264 .....LysLysGlnPhe 267
    :|||.....: |||.....: |||.....: |||.....: |||
997 AACGAC.....GATATATATGCGACAGCAAAATATCATTCG 1031
    :|||.....: |||.....: |||.....: |||.....: |||
268 AlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGlnGln 284
    :|||.....: |||.....: |||.....: |||.....: |||
1032 CAATGACAGACATCTCTCCCTAATAGATTAACAGCAACGCTTC 1081
    :|||.....: |||.....: |||.....: |||.....: |||
284 yGlnHisThrPasn.....IleThrPheGlyThr... 294
    :|||.....: |||.....: |||.....: |||.....: |||
1082 AATGTTTAAATGTTCTTATCCAGACAGCAAGAGAACTGTTATCAT 1131
    :|||.....: |||.....: |||.....: |||.....: |||
295 .....AsnSerHisIleGlySerThrAla.....ValArg 304
    :|||.....: |||.....: |||.....: |||.....: |||
1132 GCTGACAGTGGTTCACACATATGACCCAGACTGAATATGAGAAA 1181
    :|||.....: |||.....: |||.....: |||.....: |||
305 LeuAlaGlyAsnGlnLysArgAsp.....AlaAsnAsnGlyGlnAs 317
    :|||.....: |||.....: |||.....: |||.....: |||
1182 TATTCCTTATTTAGCAAGAAAGCGCAATTTGATCTTACCAACA 1231
    :|||.....: |||.....: |||.....: |||.....: |||
317 nValThrPhe.....GlnAspAsnGlyThrLeuValLeuAspGlnAsnI 332
    :|||.....: |||.....: |||.....: |||.....: |||
1232 TCAATCAAGTCTGAGAGATTAATTTCCAAAGAGATTTACGCTTCG 1281
    :|||.....: |||.....: |||.....: |||.....: |||
332 lAsnGlnGlyAlaGlyLysLeuPhePheLysGlyAspTyrThrValLys 348
    :|||.....: |||.....: |||.....: |||.....: |||
1282 CCTGAATAATACGA...ACTTGGCAAGCGCGGCTTCATATCATGTA 1328
    :|||.....: |||.....: |||.....: |||.....: |||
349 GlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAs 365
    :|||.....: |||.....: |||.....: |||.....: |||
1329 AGCAGTACCGTTACTTGGAAAGTAACGCGTGGCAACGACCCCTGT 1378
    :|||.....: |||.....: |||.....: |||.....: |||
365 pGlyLysLysValValTrpGlnValLysAsnProAsnGlnLysAspArgLeuA 382
    :|||.....: |||.....: |||.....: |||.....: |||
1379 CCAAAATCGGCAAGGACGCTGCGACGCTGCAAGCAAGGAGAAACCA 1428
    :|||.....: |||.....: |||.....: |||.....: |||
382 lAsnIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGln 398
    :|||.....: |||.....: |||.....: |||.....: |||
1429 GCGTCATACGCTGGCGAGCGATACATCATTTGGATCAGGAGCACA 1478
    :|||.....: |||.....: |||.....: |||.....: |||
399 GlyGlnLeuLysValGlyLysPglTyrThrValIleLeuAsnGlnLysAlaAs 415
    :|||.....: |||.....: |||.....: |||.....: |||
1479 CGATTAAGCAAAACAAAGCCTTTAGTGAATCGGCTGGTCAGGCGCA 1528
    :|||.....: |||.....: |||.....: |||.....: |||
415 pSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValArgGlyA 432
    :|||.....: |||.....: |||.....: |||.....: |||
1529 GGGGTACCGGCAACGTAATCCGATATATCATGTTCAACCCGCAAACTC 1578
    :|||.....: |||.....: |||.....: |||.....: |||
432 rglGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeu 448
    :|||.....: |||.....: |||.....: |||.....: |||
1579 TATTTGCGCTTTCGCGGAGCTTTGATTAACGGCATTCGCTTTC 1628
    :|||.....: |||.....: |||.....: |||.....: |||
449 TyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 465
    :|||.....: |||.....: |||.....: |||.....: |||
1629 GTTTCACCGCTATTCAAATATCCGATGAGAGGCGGATTTGTCACACACA 1678
    :|||.....: |||.....: |||.....: |||.....: |||
465 rPheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAlaHisAla 482
    :|||.....: |||.....: |||.....: |||.....: |||
1679 ATCAAGCAAGAATCCACCGTATACCTTACAGCAATATAGATTTGCT 1728
    :|||.....: |||.....: |||.....: |||.....: |||
482 snThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIle 497
    :|||.....: |||.....: |||.....: |||.....: |||
1729 ACA 1731
    :|||.....: |||.....: |||.....: |||.....: |||
498 Thr 498
  
```



```
|||||
338 leasngllylaglyleuphehnglyasntrythrvalglu 354
1279 TCGCCTGAATAATACGAAACTGGCAAGGCGGCGTTCATATCACTGA 1328
355 glyserasrpsnllvalttrpsncllyalaglyleservalagl 371
1329 AGACNATACCTTACTGTGAAAGTAAACGGCTGGCAACGACGCTGT 1378
371 uglysthrvalthrtrpysvalhhsasnproglinsersparyleua 388
1379 CCAAAATCGGCAAGCGCTGCACTTACACCCAAAGGCAAAACCAA 1428
388 lalysileglylysglythrleuilevalgluilelysglyluasnlys 404
1429 GCGTCAGCAGCGTGGCGAGGTACAGTCATTTTGATCAGCAGCACA 1478
405 glyserleuysvalglyaspglythrvalilleleuileysglinglnlaas 421
1479 CGATAAAGCAAAAACACACCTTTAGTGAAATCGGCTTGTCAGCGCA 1528
421 palasnasnlysvallysalapheserclnvalglyilevalserglya 438
1529 GGGGTACGGTGCACATGATGCGGATATCAGTTCACCCGCAAACTC 1578
438 rgsertthrvalvalleuasnaspspysglnvalaspronserserile 454
1579 TATTTCGGCTTTCGCGC 1596
455 tyrpheglyphearggly 460
seq_name: p1r2:S61333
seq_documentation_block:
IGA specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strain
N:Alternate names: IGA1 protease
C:Species: Haemophilus influenzae
A:Variety: HK284
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61333
R:Lonholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IGA1 protease in Neisseria
A:Reference number: S61314; MUID:95302961
A:Accession: S61333
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-471 <LOW>
A:Cross-references: EMBL:X82487; NID:9773205; PIDN:CAA57870.1; PID:9773206
A:Experimental source: strain HK284
C:Genetics:
A:Gene: iga
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
alignment_scores:
Quality: 677.50 Length: 520
Ratio: 2.111 Gaps: 24
Percent Similarity: 61.731 Percent Identity: 35.000
alignment_block:
US-09-303-518D-649 x S61333 ..
Align seg 1/1 to: S61333 from: 1 to: 471
202 CGGAAGATTGAGTTGACACAAAAAGGAGTGTGCGCAATC 251
|||
1 Alathrsnvalgluvalargaspysasnasnglnserleuylseral 17
252 AATGACAAAGCC...CCGATGATGATTTTCTGTGTCGTGCGCAATC 297
|||
17 aleuproasnlylleprowmetileasprheservalvalaspsvalasn 34
```

```
298 ..GCGTGGCGGCATTTGGTGGGATCATATATTTGTGACGTCGACAT 345
|||||
34 ysargileglythrleuvalasproglinttryllevalservallyshs 50
346 AAC.....GGCGCTATACAACTGTA 368
51 Alahistlyshstleasnaspphetyrpheglyhstysasn..... 64
369 TTTTGTGCGGGAAGAAATCCGATCAACATCGTTTACTTATTAATA 418
65 ..Glyhisargaspvalseraspraspaslnlys.....tyrser 78
419 TTGTGAACGGAATATATTAAAGCAGGACTAAAGCCAT.....CCT 462
78 alvalthrlnasnasnvallyspasnglnasprphtstvalasps 94
463 TATGGCGCGATATATCATATGCGCGTTGCAATTAATTTGTACAGATGC 512
95 Argleuaspsprtyrasnmetproatgleuasnlysphervalthrgluva 111
513 AGAACCTGTTGAATATGACCATATATGATGAGG...CGAATATATATCG 559
111 lalaprothrthrprothrleualaglyaspspleuclutntrytlysa 128
560 ATCAAAATATATTCCTGACCGCTGTCGATTTGGGCGCAGGCAATAT 609
128 splysglnustytrproserphevalargvalglalaglythrclnph 144
610 TGCGATCTGAT.....GAAATGAGCCATTAACCCGCA 644
145 valtyrclnlysglysertrythrvalgluysrthrtrargasnasnas 161
645 AACTTATATATCATATTCAGTGCCTATTCCTGCTGCTGGGACATA 694
161 pilelysphe...leuasplualatyrargtyralatleuglylythr 177
695 CTTTGCA.....CAAAATGATCAGTGGTGGCGACACTC 729
177 rotyrclnlylleasnlleaspproserglnserlysglyleuile 193
730 AACTTAGTAGTAGTGAATAATTAACAT..... 756
194 glypheglyaspserarglnuasnhisvalilleasphalysrthrleu 210
757 .....AGCCATATGTTTTCACCAACAGAGGCTCAT 790
210 uSerlnaspproleuthrasntryrlyvalleu..... 221
791 TTGGGAGAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAG 840
222 ..Glyaspserglyserproleuthralpheasplysglnglnaslnlys 237
841 TGCTTAATTAATGGGCTATTCGAACGGCAACCCCTAT..... 879
238 tyrvalpheilegly.....Protyrthrtrytrpal 248
880 ....ATAGAAATAAGCAAT.....GGCTCCAGCTGTCGTAAGAT 919
248 aglytyrcllylsyssertrpglnglnutrpasntletryrlylsasp. 264
920 GGTTCATGATGAATCTTTCGAGATACCCATTCAGTATTCAGCA 969
265 ..phealaspasnllelylsargaspasnlaglualaval..... 278
970 CCAGCTGAATAATGGGAATATCTTTTAAGACAGATATATATGCGCAGC 1019
279 Proheserthrserglutryrthsttrphtnasntrthrasnhsngln 295
1020 AAAAATC...AATGCCAATGATGACAAAT.....TCTCTGC 1054
295 rgluilelyasnthrasphsthrillerthrvalthrleuproseraspr 312
```

```

1055 CTAATGATTTAAAAACGACGACGCTTCATGTTTAAATGTTCTTTATCC 1104
      |||||
312 rosnarglueu..... 315
1105 GAGACAGCAGAGACCTGTTTATCATGCTGCAGGTGCTCACC...AG 1151
      |||||
316 .....Valasnhegl 319
1152 TTATGACCCAGACTGTAATATGAGAAAAATATTTCTTTATTTAGC.... 1197
      :.....:|||||:|||||:|||||:|||||:|||||
319 nglnlysglnhisleuglnthrclglnsnvalthrphespspsert 336
1198 .....GAGGAAAAAGCGGAATGTAATCTTACAGCAACATCAATCAAGGT 1242
      :.....:|||||:|||||:|||||:|||||:|||||
336 hrasnansnglylsylThrleuileuaspaspshisileasnnglyly 352
1243 GCTGGAGATTTATTTCCAAAGAGATTTACGGTCTCGCTCAAAATATA 1292
      |||||
353 Alaglylyleuphepnepelyslntrglvallylsyllysthras 369
1293 CGAA...ACTTGGCMAAGCGCGCGCTTCATATCATGTAAGACATACCG 1339
      :|||:|||||:|||||:|||||:|||||:|||||
369 paspllethrtrpvalglyglylyleasprvalalagllylylsylly 386
1340 TTAATTGAAAGTAACGCGCTGGCAACGACCGCTTCGCAAAATGCGC 1389
      |||||
386 alvaltrpelysvalhisasnproglulysasphisleualalysllely 402
1390 AAAGGACGCTGCAGCTTCAAGCCAAAGGGGAAACCAAGGCTCGATCAG 1439
      |||||
403 lysglythrleuilevalgluglylylsyllysnslslyserleu 419
1440 CGTGGGCGACGGTAACGTCAATTTTGATCAGCAGCAGCAGCATAAAGCA 1489
      |||||
419 svalglyspglythrvalvalleuylsnglnthrasmelysergly 435
1490 AAAAACAAGCCTTAGTAATTCGGCTGTGTCAGCGCAGCGGATACGGTG 1539
      :.....:|||||:|||||:|||||:|||||:|||||
436 ..gluhisalapheklaservalglyllevalserglyargserthrval 451
1540 CAACTGAATGCGGATTAATCACTTCAACCGCAGCAAACTTATTTGCGCTT 1589
      |||||
452 Valleunasnaspplysglvalalasprianseriletyrphglyph 468
1590 TCGCGCGCGCA 1599
      |||||
468 earglyglyly 471

seq_name: plf2:T00317
seq_documentation_block:
probable serine proteinase esp, extracellular - Escherichia coli plasmid p0157
N/Alternate names: putative exoprotein-precursor
C/Species: Escherichia coli
C/Date: 01-Feb-1999 #sequence-revision 01-Feb-1999 #text-change 08-Dec-2000
C/Accession: T00317; T42120
R/Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yatsudo, H.C., Kubota,
S., Shinagawa, H.
DNA Res. 5, 1-9, 1998
A/Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic
A/Reference number: Z14127; MUID:98290540
A/Accession: T00317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1300 <MAX>
A/Cross-references: EMBL:AB011549; NID:94589740; PIDN:BA31826.1; PID:93337077
A/Experimental source: strain EHEC O157:H7, substrain KMD 0509952
R/Burland, V., Shao, Y., Perna, N.T., Plunkett, G., Sofia, H.J., Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A/Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia
A/Reference number: Z22068; MUID:98391744
A/Accession: T42120
A/Status: preliminary; translated from GB/EMBL/DBJ

```

```

A/Molecule type: DNA
A/Residues: 1-1300 <BUP>
A/Cross-references: EMBL:AF074613; PIDN:AC70088.1
A/Experimental source: strain EDL933; serotype O157:H7
A/Genetics:
A:Gene: espP
A:Gene: plasmid p0157
C:Superfamily: Iga-specific metalloendopeptidase
C/Keywords: serine proteinase

```

alignment_scores:

```

Quality: 635.00 Length: 1506
Ratio: 0.863 Gaps: 67
Percent Similarity: 48.871 Percent Identity: 22.112

```

alignment_block:

```

us-09-303-518d-649 x T00317 ..

```

```

Align seg 1/1 to: T00317 from: 1 to: 1300

```

```

82 TTAGCCATATGCTGTGCTGCTGCTATCTCCCAAGCGTGGCGGACA 131
      |||||
40 LeuAlaLeuCysPhe...LeuGlyLeuLeuGlnSerSerTyrSerPheAl 55
132 CACTTATTTGCGCATCACTACCAATAC.....TATGCGCACTTG 172
      :|||
55 aserglnmetlaplleserasnphetrlylleargasptrymetaspheA 72
173 CGGAATATAAGCAAGTGTGACGCGGGCGGAAGATTTGAGGTTTAC 222
      |||||
72 laglnslsnglysllepheglnalaglyalathasnleluleval 88
223 AACCAAAAAGGCGAGTTGGTGGCAATATCAATGACAAAAGCCCGATGAT 272
      :|||
89 LysLysAspGlySerThrLeu....LysLeuProgluValIleProPheTr 103
273 TGAATTTTCTGTGCTGCTCCGTAACGCGTGGCGCATGTGGCGGATC 322
      |||||
103 oaspheserProvalAlaslnslslyserThrThrSerTleGlyGlyA 120
323 AATATATTTGAGCGTGGCAGCATAC...GGCGCTATACACGTTTGAT 369
      |||||
120 latyrserllethrAlathrHisasnThrLysasnHisHisSerValAla 136
370 TTGTGTGCGGAGAGAAATCCGATCAACATCTGTTACTTATTAAT 419
      :|||
137 ThrGlnasnTrpGlyasnser.....ThrTyrLysgl 147
420 TGTGAACGCAATATTTATAACAGAGACTAAGGCCATCTTATGGCG 469
      :|||
147 nrthasprrpAsnThr.....SerHisPro..... 155
470 GCGATTTCATATGCGCGCTTTCATTAATTTGTC..... 504
156 ..AspPheAlaValserAlaThrLeuAspSpsheValValglutrrargly 171
505 ...ACAGATCGAGAACCTGTTGAATGACCACTTATATGATGGGGGAA 551
      |||||
172 AlathrGlnGlyAlaAspIleSerLeuSerLysolnleAlaLeuGluar 188
552 ATATATGATCAAAATTAATACCTGACCGTGTTCGATTTGGGCGAGCA 601
      :|||
188 gtyr.....GlyValasnTyrLysGlyGlyLysLeuIleAlaPheA 203
602 GGCATATTTGGCATGTGATGAGATGAGCCCAATAACGC..... 642
      |||
203 rgalaglyserGlyValValserValLysasnGlyArgIleThrPro 219
643 .....GAAGTCAAT.....CA 656
      |||
220 PheasnGlnValSerTyrLysProgluMetLeuasnGlySerPheValHI 236

```

[illegible]


```

3905 TTTCAGACGC..... 3915
1234 euAlaasnglYgluThrValleuArAspAlaSerIylgluYsArgIle 1250
3916 ATCGGAGCGCAAAATCCGCCGCCGTCGTCATTCACGCATTACAGCAGC 3965
1251 LysglYgluYsAspSerArGmetLeuMetSerValIglYleuAsnAlaGl 1267
3966 ATACCGC.....GCCGGTTTCG 3982
1267 uilelYgAspAsnValArPheglYleuGluphegluYsSerAlaPheG 1284
3983 GCGGATTTCGCGATCGAACCGCACATCGCGCAACGGCTATTTCGCCAA 4032
1284 LysTyrAsnValAspAsnAlaVal..... 1292
4033 AAGCGGATTACCGCTAC 4050
1293 AsnAlaAsnPhaArgTyr 1298

```

seq_name: p1r2:B41500

seq_documentation_block:

19A-specific metalloendopeptidase (EC 3.4.24.13) type 2 - Haemophilus influenzae (fragme
N:Alternate names: Immunoglobulin A1 proteinase type 2
C:Species: Haemophilus influenzae
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Dec-2000
C:Accession: B41500
R:GrundY, F.J.; Plant, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A:Title: Localization of the cleavage site specificity determinant of Haemophilus influe
A:Reference number: A41500; M01D:90129281
A:Accession: B41500
A:Status: not compared with conceptual translation
A:Residues: 1-384 <GRU>
C:Comment: This 19A1 proteinase is classified as type 2 because it cleaves at a proline
C:Superfamily: 19A-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 400.00 Length: 426
Ratio: 1.633 Gaps: 20
Percent Similarity: 57.512 Percent Identity: 29.108

alignment_block:

US-09-303-518D-649 x B41500 ..

Align seg 1/1 to: B41500 from: 1 to: 384

```

64 GCGTTCGCGCTGCTTACTTAGCCATATGCCGTGCTTGGCATTCCTCC 113
5 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCCCTGGGCGGACACACTTATTCGGCATCACTACCAATACATC 163
21 cTyThrGluAlaAlaLeuValArGAspArpValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGAAATPAAAGCAAGTTTCAGTCGGGCGGAAAGATAT 213
38 rGAspPheAlaGluAsnLysGlyArPheSerValGlyAlaThrAsnVal 54
214 GAGGTTCACACAAAAGAGGAGTTTCGTCGCAATCAATGACAAAAGC 263
55 GluValArGAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71
264 C...CCGATGATTGATTTTCTGTGTGTCGCGTAC...GGCGTGGCG 307
71 yIlePheMetIleAspPheSerValValAspValAsnLysArgIleGlyT 88
308 CATGCGTGGCGCATATATATTGTGAGCGTGGCATAAGCGCGCTAT 357

```

```

88 hrLeuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104
358 AACAGCTGATTTT.....GTCGCGAAGAGAA 389
105 MetAsn...AspPheTyrPheGlyHisTyrAsnGlnHisArgAspValSe 120
390 TCCGATCAACATCGTTTACTTATPAAATGTGAACGATATATATA 439
120 rAspAspGluAsnLys.....TyrSerValAlaThrGlnAsnAsnValA 135
440 AACGAGGACTAAAGCCAT.....CCTTATGCGCGCGATTATCATG 483
135 snProAsnGluAsnTyrPheHisValAspLysArgLeuAspAspTyrAsnMet 151
484 CCGCGTTTCATPAAATTTTCACAGATGAGAGACCTGTGAATCAGCAG 533
152 ProArgLeuAsnLysPheValThrGluValAlaProThrThrProThrLe 168
534 TTTATGATGGG...CGGAAATATATGATCAATCAATATACCTGACC 580
168 uAlaGlyAspAspLeuGluThrTyrLysAspLysGluLysTyrLeuSerP 185
581 GTGTCGTATTCGGGCGAGCGAGCA...TATGCGCATTCGATGA 624
185 heValArGValAlaGlyArGlnLeuValTyrGluLysGlySerHis 201
625 .....GATGAGCCCAATAACCGCAAGTTCAATATCATATTCAGACG 668
202 HisValGluAspLysGlnHisGlyLysPheLysPheLysAspLeuSerAla 218
669 GATATCTTGGCGCTGTCGTCGCAATACCTTGCA.....C 703
218 aTyrTrgTyrAlaIleGlyGlyThrProTyrLysGlyIleAsnLeuAsp 235
704 AAATGATCAGCTGTGCGACAGTCACTAGTAGGAAATATAA 753
235 roSerGlnSerLysLysGlyLeuIleGlyPheGlyAspSerArgGluAsp 251
754 CAT.....AGCCGATA 764
252 HisValIleAsnSerLysThrLeuLeuSerGlnAspProLeuThrAsnTyr 268
765 TGGTTTTCACCAACAGAGGCTCATTTGGCGAGAGTGCACCAATC 814
268 rGlyValLeu.....GlyAspSerGlySerProLeu 279
815 TTATCTATGATGCCCAAAAGCAAGTGTAAATTAATGGGATATTCGAA 864
279 heAlaPheAspLysGlnGlnAsnLysTrpPheIleGly..... 292
865 ACGGCAACCCCTAT.....ATAGGAAAACGATATGCTT 899
293 .....ProTyrTrpTyrTTPAlaGlyTyrGlyLysLysSer..... 304
900 CCAGCTGGTTGTAAAGATGG...TTCATGATGAATTTTGGTGGAG 946
305 .....ThrGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 318
947 ATACCCATTCAGTATTCTACGACACGCTCAAAATGGAAATACCTCTT 996
318 spVal.....Leu 320
997 AACGACATATATATGSCACAGAAAATCATGCGCAACATGACACAA 1046
321 AsnLysAspSerAlaGlyLeuLeuLysGlyHisThrGlnTyrAsnTrpH 337
1047 TTCCTGCTCAAT...AGATTAAACACAGAACCGTTCAATTGTTAATG 1093
337 rSerAsnGlyAsnThrSerMetIleSerAsnGlySerGlnLeuLeuGlyY 354
1094 TTTCCTTATCGGACAGACAGAGAACCTGTTATATCATCTCAGAGTG 1143
354 aLAsnLeuPheAspAsnSerLys..... 361

```

1144 GTCAACAGTATGACCCAGACTGATTAATGAGAAATATTTCTGTTAT 1193
 : : : : :
 362 ...HISTHASNIRGLULYSALASNTYRGLYLSERVALTHRPH... 376
 : : : : :
 1194 TGACGAGAAAGGCGCAATGTACTT 1221
 : : : : :
 377 ...GlnGlyAsnGlyThrLeuThrleu 384
 seq_name: p1r2:C91068

seq_documentation_block:
 hypothetical protein Ecs3515 [Imported] - Escherichia coli (strain O157:H7, substrain R1
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C91068
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C91068
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1571 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA36938.1; PID:q1362986; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 050952
 C:Genetics:
 A:Gene: Ecs3515

alignment_scores:
 Quality: 367.00 Length: 1689
 Ratio: 0.470 Gaps: 88
 Percent Similarity: 46.240 Percent Identity: 20.130

alignment_block:
 US-09-303-518D-649 x C91068

Align seg 1/1 to: C91068 from: 1 to: 1571

268 ATGATGATTTTCTGTCGTCGCGTAACGCGCGCGCATGTCGG 317
 : : : : :
 60 LeuLeSerAlaLeuIleTyrAlaProProGlyMetAlaAlaPheThr 76
 : : : : :
 318 CGATCAATATATTGACCGTCGGACATAC.....GGCGGCTATA 358
 : : : : :
 76 OASP.....ValIleGlyValAlaAsnAspGluThrValAspLysSerG 91
 : : : : :
 359 ACAACGTTGATTTTGTGCGAGAGAGAAATCCCGATCAACATCGTTT 408
 : : : : :
 91 LnaArgValAsp.....GluArgGlyThrThrAsnAsnThrHis..... 103
 : : : : :
 409 ACTTATTAATTTGTAACGCAATATATAAGCAGGACGACTAAAGGCA 458
 : : : : :
 104IleIleAsnHisGlyGlnGlnAsnVal..... 112
 : : : : :
 459 TCCTTATGCGCGCGCATTTATCATATGCCGCTTGCATAATTTGTCACAG 508
 : : : : :
 113 ...TyrGlyGly.....ValSerA 118
 : : : : :
 509 ATGCAAGACCTTGTGAATGACCACTTATATGAT...GGCGG..... 549
 : : : : :
 118 snGlySerLeuIleGlnSerGlyGlyGlnAspValGlyArgHisAsn 134
 : : : : :
 550 AATATATGATCAAAATATATACCTGACGCTTTCGATTTGGGCGAG 599
 : : : : :
 135 AsnThrValGlyGlnSerAsn.....AsnThrThrIleAsnGlyG 148
 : : : : :
 600 CAGGCAATATTTGGCGATCGATGATGAT..... 627
 : : : : :
 148 yArgGln.....SerIleHisAspGlyGlyIleSerThrGlyThrI 162

628GAGCCCAATACCGGGAAGTTCATATCATATTATTCAGTGCCTAT 672
 : : : : :
 162 LeileGlySerGlyAsnGlnAspValIleTyrGlyGlyIleSerAsnGly 178
 : : : : :
 673 TCTTGCGCTGTTGTCGCAATACCTTTGCACAAAATGATCAGTGTGCG 722
 : : : : :
 179 ThrThrIleGlyGlyAlaSerArgValGluGlyGlySerAlaAsnGly 195
 : : : : :
 723 CACA...GTCACTTAGTACTGAAAAATTAACATTAACCCATATGCTT 769
 : : : : :
 195 yThrLeuIleAspGlyGlySerGlnIleValValGlnIleGlyHisAla 211
 : : : : :
 770 TTTTACCAACAGAGGCTCATTTGGCGACAGTGCCTACCAATGTTATC 819
 : : : : :
 212AspGlyThrThrIleAsnLysSerGlySerGlnAspValVal 225
 : : : : :
 820 TATGATGCCCAAAAGCAAAAGTGTATTAATTAATGCGATTCACACGG 869
 : : : : :
 226 GlnGlySerLeuValThrAsnThrThrIleAsnGly.....G 238
 : : : : :
 870 CAACCCCTATATAGGA.....AAAAACA 892
 : : : : :
 238 yArgGlnIleValGluGlnSerThrValGluThrThrThrIleLysAsnG 255
 : : : : :
 893 ATGCTTCACAGCTGTCGTAAAGATGTCCTATGATGAATCTTTCCT 942
 : : : : :
 255 yGlyGluGlnIleArgValTyrGlnSerArgAlaLeuAspThrThrIleGlu 271
 : : : : :
 943 GGAGATACCCATTCAGTATTCAGAACACAGCTCAAAATGGGAA..... 987
 : : : : :
 272 GlyGlyThrGlnSerLeu.....AsnSerLysSerThr 282
 : : : : :
 987 987
 : : : : :
 282 LalaLysAsnThrGlnIleTyrSerGlyGlyThrGlnIleIleAspAsnT 299
 : : : : :
 988TACTCTTTAAC.....GACGAT 1005
 : : : : :
 299 hrSerSerAspValIleGluValTyrSerGlyGlyValLeuAspVal 315
 : : : : :
 1006 AATATGCGACAGAAATCAATGCGCAACATGACACACATTCCTGCGC 1055
 : : : : :
 316 SerGlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaIleLeuGly 332
 : : : : :
 1056 TAATAGATTAAACACAGAACCTTCATTTGTTATGTTCTTTATCGG 1105
 : : : : :
 332 strAsnThrAsnGlyThrThrValSerGlyThrAsn.....SerG 346
 : : : : :
 1106 AGACAGCAAGAAACCTGTTATCATGCTGCA..... 1137
 : : : : :
 346 LuGlyAlaPheSerIleHisAsnHisValAlaAspAsnValLeuLeuGlu 362
 : : : : :
 1138 ...GGTGT.....GTCAACAGTTAT..... 1155
 : : : : :
 363 AsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerAlaAsnLysThrI 379
 : : : : :
 1156C 1156
 : : : : :
 379 eileLysAspGlyGlyThrMetSerValLeuThrAsnAlaLysAlaAspA 396
 : : : : :
 1157 GACCCAGACGATTAATGGA..... 1176
 : : : : :
 396 IatThrArgIleAspAsnGlyGlyValMetAspValAlaGlyAsnAlaThr 412
 : : : : :
 1177GAAATATTCCTTTATTCAGCAGAG 1202
 : : : : :
 413 AsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnGlyGlyLeal 429
 : : : : :
 1203 AAAGGCGAATTTACTTACCAAGACATCAATCAAGT..... 1242
 : : : : :
 429 aThrGly.....ThrAsnIleAsnSerGlyThrGlnAsnI 441
 : : : : :
 1243GCTGAGATTAATTTCCAAAGAGATTTTACGTCGCCCTGAA 1287

441 LelysSerGlyGly.....LysAlaAspThrThrIleIleSerSer 454
1288 AATTAAC.....GAACCTTGGCAAGCGCGCGCTTCAT 1322
455 GlySerArgGlnValValGluLysAspGlyThrAlaIleGlySerAsnII 471
1323 CAGTGAAGACAGTACCGCTTACTTGGAAGTAAACGGCGGTGCAAAACGAC. 1371
471 eSerAlaGlyGlySerIleuIleValIlyThrGlyGlyIleAlaHisGly 488
1372CGCTGTCCAAATCGCAAAAGCGACG 1398
488 AlaSngInGluThrGlySerAlaLeuValAlaAsnThrGlyAlaGlyThr 504
1399CTGCAGCTTCAAGCCAAAGG..... 1419
505 AspIleGluGlyTyriAsnLysLeuSerHisPheThrIleThrGlyGly 521
1420GAAACCAAGGCTCGATCAGCGGTGGCGACG 1450
521 uAlaAsnTyiValValLeuGlnAsnThrGlyGluLeuThrValAlaAla 538
1451 GTACAGTCATTTGGATCAGCAGCAGATAAAGCAA..... 1491
538 ysthSerAlaLysAsnThrThrIleAspAlaGlyLysLeuIleVal 554
1492AAACAAACCTTTAGTGAATCGGCTTGTGCAGCGCAGGGTACGCT 1538
555 GlnLysGlnAlaLysThrAspSerThrArgLeuAsnAsnGlyLysAla 571
1539 GCACAGCAATGCCGAT.....AATCAGTTCACCCCG 1570
571 uGluValGlnAspGlyGlyLysAlaLysHisValGlnGlnInserylg 588
1571 ACAACTCTATTTCCGCTTCCGGCGGACGTTG.....GATTAAAC 1614
588 LysAlaLeuIleAlaSerThrThrSerGlyThrLeuIleGluGlyThrAsn 604
1615 GGGCATTCGCTTTCGTTCCACCGTATTCAAATACCGATGAAGGGCGCAT 1664
605 SerTyiGlyAspAlaPheTyi...IleArgAsnSerGlnAlaLysAsnVa 620
1665 GATTGTCAACCAACATCAAGACAAAGATCCACCGTTACCATTAAGCA 1714
620 ValLeu.....GlnAsnAlaGlySerLeuThrValValThrGlys 634
1715 ATAAAGATTTGCTACACCGGCAATTAACACAGCTTGATGCAAAAAA 1764
634 eArgAlaValAspThrIleIleAsnAlaAsn.....GlyLysPhe 647
1765 GAATTTGCTTACACAGGTTGGTTGGCGAAGAAAT.....ACGACCAA 1808
648 AspVal.....TyiGly...LysAspValGlyThrValLe 658
1809 AACGAACGGCGGCTCAACCTTGTATACAGCCGCCCGCAGAAAGCCCA 1858
658 uAsnSerAlaGlyThrGlnThrIleIleValSerAlaThrSerAspLys 675
1859 CCGCTCTCTCTTCGCGGCAACAAATTAACGGCAGCATCAGCAACA 1908
675 LaAsnIleLysGlyLysGlnThrValTyiGlyLeuAlaThrGlnLys 691
1909 AAC.....GCAAACTGTTTTCAGCGGCGACAGCAACCGCAAGC 1949
692 AsnIleGlnSerGlyGlnGlnIleValAspGlySerThrGluLys.. 707
1950 CTACAATCATTTAAACGAC..... 1968
708ThrHisIleAsnGlyLysGlnThrValGlnAsnTyiGlyLys 723
1969CATTTGGTGCAGAAAGAGGCG 1989

723 IAlaAsnThrAspIleValSerGlyLeuGlnIleMetAlaAsnGly 739
1990 ATTCTCGCGGGAATTCGTG...TCGCACACGACGTGATCAACCGC 2036
740 ThrAlaGlnGlySerIleIleAsnGlnGlySerGlnIleValAsnGln 756
2037 ATTTAAACGGGAAACTTCCAATTAAGCGGACAGCGGTGTTCC 2086
756 yGlyLeuAlaGlnAsnSerValLeuAsnAspGlyGlyThrLeuAspVal 773
2087 GCATGTTTCCAAAGGCAAGGCGATGGCATTTGACCAATACGCCCA 2136
773 rGlnLysGlySerAlaThrGly.....IleGlnInserylgln 786
2137 GCAGTTTGTGTGCGACCGCATCAAGCCACACAAATCTGTACGTTG 2186
787 GlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArgAl 803
2187 GGACTGACGGGTGACCAATTTGTGCAAAA..... 2220
803 asp.....GlyValAlaPheSerIleGluGlnGlyAlaAlaAsnAsnI 818
2221ACCATTCAGACAT..... 2235
818 leLeuLeuAlaAsnGlyGlyValLeuThrValGlnSerAspThrSer 834
2236AAAGTATTGC 2246
835 AspLysThrGlnValAsnThrGlyLysArgGluIleValLysThrLysAl 851
2247 TTCAATGACATAAGCCGATCAGCGGCAATGTGATCTT..... 2286
851 eThrAlaThrGlyThrThrLeuThrGlyGlnGlnIleValAlaGly 868
2287 ..GCCGATCAGCTCATTTAATCTCACAGGCTTGCACACCTCAAGCGC 2334
868 AlaAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnThr..... 881
2335 AATCTTAGTCAAAATGGCGATACA...CGTTATACAGTCAACCCAAAGC 2381
882 ..ValSerAlaAsnGlyGlnAlaIleLysThrThrIleAsnGlyGly 897
2382 CACC.....CAAAAGCGCAACCTTAGC...CTGTGGGCAATG 2416
897 YThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValAlaAsn 914
2417 CCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGCTTCG 2466
914 eArgAlaAlaLeuGlnThrSerThrAlaAsnGly...IleGluIleSer 929
2467 GGCAT.....GCTCATTTAATCTAAGCGACACCGCTACAAA 2507
930 GlnThrHisGlnTyiGlyThrPheSerIleSerGlyAsnLeuAlaThrAs 946
2508 CGGCACTGTGACGCTTCCGGCAACGCTAAGCAACGTAAGCATTCG 2557
946 nMetLeuLeuGlnAsnGlyLysAsnLeuValIleAlaGlyThrGln 963
2558 CACTCAACGGTAATGCTCCCTAGCCGATTAAGCAGATTTCCATTTGAA 2607
963 IaArgAspSerThrValGlyLysGly...AlaMetGlnAsnGlnGly 978
2608 AGCAGCGCTTTACCGGCAAAATCAGCGGCGGCAAGATACGGCATTA 2657
979 GlnAspSerAlaThrLysValAsnSerGlyGlnTyiThrLeuGlnYar 995
2658 CTTAAAGACAGCAATGACGCTGCGTACGACGAGATTA..... 2700
995 gSerLysAspGluPheGlnAlaLeuAlaArgAlaGlnAspLeuGlnVal 1012
2700 2700
1012 IaGlyGlyThrAlaIleValTyiAlaGlyThrLeuAlaAspAlaSerVal 1028

2701GCAATTTAAACCTT.....GACAGCCGAC 2726
1029 SerGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValTh 1045
2727 C...ATTACCTCAATTCGGCTATCGC...CAGCATGGCGGCGGCGC 2770
1045 rProValLysLeuGluGlyAlaIleArgIleThrAspSerAlaThrLeu 1062
2771 AAACCGCGACGTGGCAGACATGCGCGCGCGCGCTTGGCGCGTGGCGC 2820
1062 hrIleGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSer 1078
2821 CGT.....TCCCTATTA..... 2823
1079 ArgGlySerValTyrLeuAsnSerAsnAsnSerCysAlaGlyThrSerAs 1095
2824TCCCTATTA..... 2832
1095 ncysGlyTyrAlaValAsnSerLeuLeuAsnAspGlyAsnValTyrL 1112
2833TCGTTACACCGCCCACTTGGTAGATCCGTTTCAACACG 2874
1112 euserAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTyrAsnThr 1128
2875 CTGACGGTAAACGGCAATGGAACGGTCAGGGAACATTCGCTTATGTC 2924
1129 LeuThrThrAsn...GluLeuSerGlySerGlyAsnPhetTyrLeuHisTh 1144
2925 GGAACCTCTGCGTACCGCACGACGCAAAATGMACTGGCGGAAAGTTCG 2974
1144 rAsnValAlaGlySerArgGlyAspGlnLeuValAlaAsnAsnAlaT 1161
2975 AAGCACTTACACCTGGCGGTCAACATACCGCAACGAACTGGCAGC 3024
1161 hrGlyAsnPhetLysIlePheValGlnAspThrGlyValSerProGlnSer 1177
3025 CTCGAACATTTAGCGTAGTGAAGGAAAGCAAAACCGCTGTCCGA 3074
1178 AspaAlaMetThrLeuVal.....LysThrGlyGlyGly 1189
3075 AAACCTTAATTTACCTGCAAAACGAA.....CAGTCGATCCGCGC 3118
1189 yAspAlaSerPheSerLeuGlyAsnThrGlyLysPheValAspLeuGlyT 1206
3119 CGTGGCGTTACCACTCATCCGCAAGACGCG.....GAGTTCGCGCTG 3162
1206 hrTyrGlyTyr...ValLeuLysSerAspGlyAsnSerAsnThrAsnLeu 1221
3163 CATTAATCCGTCGTAAGAACAGAGCTTTCGACAAACTGGCAAGCGCA 3212
1222 ThrAsnAspValLysPro...AsnProAspProAsnProAsnProA 1238
3213 ACCCAAAAACAGCGCGGAAAAAGACACGCGCAAAAGCTTGACGCGCTGA 3262
1238 snProLys..... 1240
3263 TTGGCGCGCGCGCGATGCGCTGCAAAAAGACAGAAAGCTTGGCGAACG 3312
1241ProAsp... 1242
3313 GCCCGGACAGCAGGCGGAAAAATGCGCATTTATGCAAGCGGAGAGAGA 3362
1242 1242
3363 GAAAAAACGGTGCAAGCGGATTAAGACACCGCTTGGGAAACAGCGCG 3412
1242 1242
3413 AAGCGGAACCGCGCGCTACACGCGCTTCCCGCGCGCGCGCGCGC 3462
1243ProLysProAsp.ProLysProAspProL 1252

3463 CGCGGATTTGCCCAACTGCAACCCCAACCGGACGCCCAACCGCAGC 3512
1252 yPro...Asp...ProThrProGlnProThrProThrProValProGlyL 1267
3513 GCACCTGATCAGCCGTTATGCCAATAGCGGTTTGANGAATTTCCGCCA 3562
1267 sarGlyLeuPro.....SerThrAlaAlaValLeuAsnMetAlaAlaT 1282
3563 CGCTCAACAGCGTTTC.....GCCGTACAGACGAATTA 3597
1282 hrLeuProLeuValPheAspAlaGluLeuAsnSerIleArgGluArgLeu 1298
3598 GACCGCGTATTTGCCGAAGACCGCGCAGCGCGCTTGGACAGCGGCAT 3647
1299 AsnIleMetLysAlaSerProHisAsnAsnValTyr...GlyAlaThr 1314
3648 CCGGACACCAAAACACTACCGTTCCAGATTTCCGCGCTTACCGCCAC 3697
1314 rTyrAsnThrArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGlu 1331
3698 AAACCGACCTGCGCCAAATCGTATGCAGAAAACCTC..... 3735
1331 InThr.....LeuThrGlyMetThrValGlyIleAspSerProAsn 1344
3736GCGACGCGCGCGCTCGCGATCCGTTTGGCAACACG 3773
1345 AspIleProGluGlyIleAlaThrLeuGlyAlaPheMetGlyTyrSerH 1361
3774 GACCGAAAACACTTCGACGACGCG..... 3798
1361 serHisIleGlyPheAspArgGlyGlyHisGlySerValGlySerTyr 1378
3799 ..ATCGCAACTCGGACGCGCTTGGCCACGCGCGCTTTC..... 3837
1378 erLeuGlyGlyTyrAlaSerTyrGlnHisGlySerGlyPheTyrLeuAsp 1394
3838 GGGCAATAGCGCATCGACAGCTTC.....TCATCGCATCGACGCG 3878
1395 GlyValValLysLeuAsnArgPheGluSerAsnValAlaGlyLysMetSe 1411
3879 GGGCGCGGTTTATGACGCGCGACGCTT...TCAGACGCGATCGAGCA 3925
1411 rSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyLeuGlyGly 1428
3926 AAATCCGCGCGCGCTGTCATTCAGGCATTCAGGCATTCAGGCAGCGCGC 3975
1428 Is.....IleGlyThrGlyMetArgPhe 1435
3976 GGTTCGGCGGATTCGCAATCGAACCGCAATCGCGCAACGCGTATTT 4025
1436 ThrAspGlyAsnThrAsnLeuThrProTyrAlaSerLeuThrGlyPheTh 1452
4026 CGTCAAAAACGCGATTACCGC.....TACGAAAACGTCATA 4063
1452 rAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGlySerTyr 1469
4064 TCGCCACACCGCGCTTGCATTCACCGCTACCGC...GGGCGCATTAAG 4110
1469 alaSpThrArgSerIle.....TyrArgGluLeuGlyAlaThr 1481
4111 GCAGATTATTCATTAACCGCGCGCAACACATTTCCATCAGCGCTATTT 4160
1482 LeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGluProThrPhe 1498
4161 GAGCGTGTCTATACCGATGCGCGCTTGGCAAAAGTCCGACACGCGTCA 4210
1498 ULysAlaAlaValArgGlyGluPheValAspAspAsnArgValLysVal 1515
4211 ATACC..... 4215
1515 snAsnAspGlyAsnPhenValAsnAspLeuSerGlyArgGlyIleTyr 1531
4216 ...GCCGATTTGGCTCAGATTTCGCAAAACCGCGAGTGGCAATGGGG 4262

1532 GlnAlaGlyLeuYsaLaserPheSerThrLeuSerGlyHisLeuGI 1548
 4263 CGTAAAGCGCCGAATCAAAAGTTTCAGCGTGTCCCTCCAGCGTCCGCC 4312
 1548 yVal.....GlyTyrSer.....HisGlyAlaGly 1557
 4313 CCAAGGCCCG 4323
 1557 algiSerPro 1560

seq_name: pir2:D85912

seq_documentation_block:
 hypothetical protein ypJ (imported) - Escherichia coli (strain O157:H7, substrain EDL93)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: D85912
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: D85912
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1528 <STO>
 A:Cross-references: GB:AE005174; NID:q12517083; PIND:AA657760.1; GSPDB:GN00145; UWGP:Z35
 A:Experimental source: strain O157:H7, substrain EDL933
 A:Genes: ypJ

alignment_scores:
 Quality: 366.00 Length: 1689
 Ratio: 0.468 Gaps: 89
 Percent Similarity: 46.300 Percent Identity: 20.130

alignment_block:

US-09-303-518D-649 x D85912 ..

Align seg 1/1 to: D85912 from: 1 to: 1528

268 ATGATGATTTTCTGTGTCGCGCTAACGCGGTCGCGCATGTGGG 317
 17 LeuIleSerAlaLeuIleTyrAlaProGlyMetAlaAlaPheThrPr 33
 318 CGATCAATATATTGTAGCGTGCACATAAC.....GGCGGTATA 358
 33 oasp.....ValIleGlyValAlaAsnSpLuhThrValaspGlySerG 48
 359 ACAACGTGATTTTGTGCGGGAAGAAATCCGATCAACATCGTTT 408
 48 LmrYValasp.....GluArgGlyThrThrAsnAsnThrHis..... 60
 409 ACTTAATAATGTGAAAGCAATATATAAGCAGGACATAAGGCCA 458
 61IleIleAsnHisGlyGlnGlnAsnVal..... 69
 459 TCCTATATGGCGCGATTCATATATCCGCTTGCATAAATTTGCACAG 508
 70TyrGlyGly.....ValSerA 75
 509 ATGCAGACCTGTGAATGACCACTTATATGAT...GGCGG..... 549
 75 snGlySerLeuIleGlnSerGlyTyrGlnAspValGlyArgHisAsn 91
 550 AAATATATGATCAAAATATATACCTGACGCTGTCTATTTGGGCGAG 599
 92 AsnTyrValGlyGlnSerAsn.....AsnThrThrIleAsnGlyGI 105
 600 CAGGCAATATTGGCATCTGATGAAGAT..... 627

105 yArgGln.....SerIleHisAspGlyLeuIleSerThrGlyThrI 119
 628GAGCCCAATAAACCGCAAGATTTCATATCATATTTGCAAGTCGAT 672
 119 LeuIleGlnSerGlyAsnGlnAspValTyrLeuGlyGlyIleSerAsnGly 135
 673 TCTGGCTGCTGTGGCAATACCTTTGCACAAATGATCAGGTGTGG 722
 136 ThrThrIleLeuGlyGlyAlaSerArgValGlnGlyGlySerAlaAsnGI 152
 723 CACA...GTCACTTAGTAGTGAATAAAATTAACATAGCCCATGTGCT 769
 152 yThrIleuIleAspGlyGlySerGlnIleValIleValGlnGlyHisAla 168
 770 TTTTACCAACAGAGCGCTCATTTGGCGACAGTGCACCAATGTTTATC 819
 169AspGlyThrThrIleAsnGlySerGlySerGlnAspValAl 182
 820 TATGATGCCCAAAACCAAAAGCTTATTAATGCGGTATTCACCAACGG 869
 183 GlnGlySerLeuAlaThrAsnThrThrIleAsnGly.....GI 195
 870 CAACCCCTATATATAGA.....AAAGCA 892
 195 yArgGlnTyrValGlnGlnSerThrValGlnThrThrIleLeuAsnG 212
 893 ATGCTTCACAGCTGTTGCTTAAGATGTTCTATGATGAATCTTTGCT 942
 212 yGlyGlnGlnLmrYValTyrGlnSerArgAlaLeuAspThrThrIleGlu 228
 943 GGAGATACCCCATTCAGTATTCACGACACCGTCAAAATGGGAAA..... 987
 229 GlyGlyThrGlnSerLeu.....AsnSerLysSerTh 239
 987 987
 239 rAlaLysAsnThrGlnIleTyrSerGlyThrGlnIleIleAspAsnT 256
 988TACTCTTTTAAAC.....GACGAT 1005
 256 hrSerSerSerAspValIleGluValTyrSerGlyGlyValLeuAspVal 272
 1006 AATATGCGCAGAGAAAATCAATGCCAATACATCAACATCTCTGCC 1055
 273 SerGlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaIleLeu 289
 1056 TAATAGATTAAACACGAAACCGTTCAATGTTTAAATGTTTCTTATCG 1105
 289 sThrAsnThrAsnGlyThrThrValSerGlyThrAsn.....Serg 303
 1106 AGACAGCAGAGAACCTGTTATTCATGCTGCA..... 1137
 303 LngIyAlaPheSerIleHisAsnHisValAlaAspAsnValLeuGlu 319
 1138 ...GGTGGT.....GTCAACAGTTAT..... 1155
 320 AsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerAlaAsnLysThrII 336
 1156C 1156
 336 eIleLysAspLysGlyThrMetSerValLeuThrAsnAlaLysAlaAspA 353
 1157 GACCCAGACTGAATATATGA..... 1176
 353 LmrYArgIleAspAsnGlyGlyValMetAspValAlaGlyAsnAlaThr 369
 1177GAAATATTTCTTTATTTGACAGAGG 1202
 370 AsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnAsnTyrGlyLeaI 386
 1203 AAAAGCGAATGATGATCTTACGACCAACATCATCAAGT..... 1242
 386 aThrGly.....ThrAsnIleAsnSerGlyThrGlnAsnI 398

1243GCTGAGATATATATTCACAGAGATTTTACGCTCGCTCGTAA 1287
1288 AATTAAC.....GAACTTGCAAGCGCGCGCTTCAAT 1322
1289 leysersglygly.....LysAlaAspThrThrIleIleSerSer 411
412 GlySerAlaGlnValValGlnLysAspGlyThrAlaIleGlySerAsnIle 428
1323 CAGTGAAGACAGTACCGCTTACTTGAAAGTAAACGGCGGTGCAACAGC 1371
428 eSerAlaGlyGlySerIleValIleValIleValIleValIleValIle 445
1372CGCTGTCACAAATCGGCAAGGCAAGC 1398
445 AlaSngInIleThrGlySerAlaLeuValAlaAsnThrGlyAlaGlyThr 461
1399CTGCAGCTTCAGCGCAAGG..... 1419
462 AspIleGlnGlyTyraSngLysLeuSerHisPheThrIleThrIleGlyGly 478
1420GAAACCAAGCTCGATCAGCTGGCGCAGC 1450
478 uAlaAsnTyraValValLeuGlnAsnThrGlyIleuThrValAlaAla 495
1451 GTACAGTCAATTTGATCAGCAGCAGCATTAAGGCAA..... 1491
495 ystHserAlaLysAsnThrThrIleAspAlaGlyGlyLysLeuIleVal 511
1492AACAAGCCTTTAGTGAATCGCTGGTTCAGCGCGGAGGTACGCT 1538
512 GlnLysGlnAlaLysThrAspSerThrArgLysAsnSngLysValle 528
1539 GCACTGAATCGCGAT.....AATGATTCACACCGC 1570
528 uGlnValGlnAspGlyGlyGlnAlaLysHisValGlnGlnIleSerGlyG 545
1571 ACAAGATATTCGCTTCGCGTTCGCGGAGCGTTCG.....GATTTAAG 1614
545 lYAlaLeuIleAlaSerThrThrSerGlyThrIleuIleGlnGlyThrAsn 561
1615 GGGCATTCGCTTCGCTTCGCGGATTCGCAATTAACGATGAAGGCGCAT 1664
562 SerTyraGlyAspAlaPheTyra.....IleArgAsnSerGlnAlaLysAsnVal 577
1665 GATTTGCAACCAATCAAGCAAGAAATCCACCTTACCTTACAGGCA 1714
577 lValLeu.....GluAsnAlaGlySerLeuThrValAlaIleThrGlys 591
1715 ATAAGATATTCGCTTCACACCGCAATTAACACAGCTTGATGCAAAAA 1764
591 eArgValAlaValAspThrIleIleAsnAlaAsn.....GlyLysMet 604
1765 GAATTCGCTTCACACCGTTCGCGGAGAAAT.....ACGACCA 1808
605 AspVal.....TyraGly.....LysAspValGlyThrValle 615
1809 AACGAACGGCGGCTCAACCTGTTTACAGCGCGCGGAGAAAGCGCA 1858
615 uAsnSerAlaGlyThrGlnThrIleTyraLysSerAlaThrSerAspLysA 632
1859 CCCTGCTGCTTCGCGGAGAAATTAACGCAACATCAACGCAAA 1908
632 lAsnIleLysGlyGlyGlnThrValTyraGlyLeuAlaThrGlnAla 648
1909 AAC.....GCAAACTGTTTTCACGCGCAGACCAACCGCAGC 1949
649 AsnIleGlnSerGlyGlnIleValAlaAspGlyGlySerThrGlnLys.. 664
1950 CTCAATCAATTAACGAC..... 1968
665ThrHisIleAsnGlyGlyThrGlnThrValGlnAsnTyraGlyLysA 680

1969CATTTGCGCAAAAGAGCGC 1989
680 lAlleAsnThrAspIleValSerGlyLeuGlnGlnIleMetAlaAsnGly 696
1990 ATTCTCGCGGGAATCGTG...TGGGACAAAGACTGGATCAACCGCAC 2036
697 ThrAlaGlnGlySerIleIleAsnGlyGlySerGlnIleValAsnGly 713
2037 ATTTAAGGGGAAACTTCCAATTAAGCGGACAGCGGCTGTTCC 2086
713 yGlyLeuAlaGlnAsnSerValLeuAsnAspGlyGlyThrLeuAspValA 730
2087 GCAATGTTCCAAAGTGAAGCGATTTGCAATTCAGCAATCAGCCCAA 2136
730 rGlnLysGlySerAlaThrGly.....IleGlnIleSerSerGln 743
2137 GCAGTTTTGTTGTCGACCGCATCAACACCAACATCTGACAGCTTC 2186
744 GlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArgAl 760
2187 GCACTGAGCGGCTGCAAAATTTGTCGAAAA..... 2220
760 asp.....GlyValAlaPheSerIleGlnGlnGlyAlaAlaAsnAsnI 775
2221ACCATTAACGACAT..... 2235
775 lLeuLeuAlaAsnGlyGlyValLeuThrValGlnLysAspThrSerSer 791
2236AAAGTATTCG 2246
792 AspLysThrGlnValAsnThrGlyGlyArgGlnIleValLysThrLysAl 808
2247 TTCATTGACTAAGACGACATCAGCGCAATGTCATCTT..... 2286
808 aThrAlaThrGlyThrThrLeuThrGlyGlyGlnIleValGlnGly 825
2287 .GCCATCAGCTCATTTAATCTCAGAGGCTTCGACACACCAAGCGC 2334
825 aAlaAsnGlnThrThrIleAsnAspGlyGlyIleGlnThr..... 838
2335 AATCTGATGCAATGCGCATACA...CGTTATACAGTACGACCAACGC 2381
839 ...ValSerAlaAsnGlyGlnAlaIleLysThrThrIleAsnGlnGly 854
2382 CACC.....CAAAACGCAACCTTAC...CTCGTGGCAATG 2416
854 yThrLeuThrValAsnAspSngLysAlaThrAspIleValGlnAsn 871
2417 CCCAAGCAACATTTAATCAAGCCACATTAAGCGCAACATCGGCTTCG 2466
871 eArgAlaAlaLeuGlnThrSerThrAlaAsnGly...IleGlnIleSer 886
2467 GGCAT.....GCTTCATTTAATCTAAGGACCGCCGCTACAAA 2507
887 GlyThrHisGlnTyraGlyThrPheSerIleSerGlyAsnLeuAlaThrAs 903
2508 CGGCACTGACGCTTCGCGCAACGCTAAGGCAACGTAAGCATTCG 2557
903 nMetLeuLeuGlnAsnGlyGlyAsnLeuValIleuAlaGlyThrGlnA 920
2558 CACTCAACGGTAATGCTCCCTAGCCGATTAAGCAGATTCATTTGAA 2607
920 lAaArgAspSerThrValGlyLysGlyLys...AlaMetGlnAsnGlnGly 935
2608 AGCAGCGCTTACCGGACAAATCAGCGCGGCAAGATACGGCATTA 2657
936 GlnAspSerAlaThrLysValAsnSerGlyGlyLysThrLeuGlyArg 952
2658 CTAAAGAC.....ACGAATGACAGCTGC 2683
952 gSerLysAspGlnPheGlnAla***AlaArgAlaGlnAspLeuGlnValA 969
2684 CGTCAGCGACGGAATTA..... 2700

```
969 1aglyGlyThrAlaLeuValTyrAlaGlyThrLeuAlaAspAlaSerVal 985
2701 .....GGCAATTAAACCTT.....GACACAGCCAC 2726
986 SerGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValTh 1002
2727 C...ATTACACTCAATTCCGCTATCCG...CACATGCGGACAGGCGC 2770
1002 rProValIysLeuGluGlyAlaIleArgIleThrAspSerIaThrLeu 1019
2771 AAACGGGAGTGGCAGATGCGCGCGCGCGCTGCGCGCTTCGCGC 2820
1019 hrIleGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSer 1035
2821 CGT..... 2823
1036 ArgGlySerValThrPleuAsnSerAsnAsnSerCysAlaGlyThrSerAs 1052
2824 .....TCCCTATT..... 2832
1052 nCysGluTyrArgValAsnSerLeuLeuAsnAspGlyAsnValTyrL 1069
2833 .....TCCGTACACGCCCAACTCCGCTAGACATCCGTTTCAACAG 2874
1069 euserAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTyrAsnThr 1085
2875 CTGACGGTAAACGGCAATTGAACGGTCAAGAAACATTCGCTTTATGTC 2924
1086 LeuThrThrAsn...GluLeuSerGlySerGlyAsnPhenylLeuIsth 1101
2925 GGAACCTTGGGCTACCGCAGCCACAAATGAACTGGCGGAAAGTTCCG 2974
1101 rAsnValAlaGlySerArgIlyAspGlnLeuValAlaAsnAsnAlaAT 1118
2975 AAGGCACTTACACCTTGGCGGTACACAAATACGCAACGAACCTGGAAC 3024
1118 hrGlyAsnPhenylIlePheValGlnAspThrGlyValSerProGlnSer 1134
3025 CTGGAACAATTGACGGTAGTGAAGGAAAAAGACAAACCGCTGTCCGA 3074
1135 AspAspAlaMetThrLeuVal.....LysThrGlyLysL 1146
3075 AAACCTTAATTCACCCGTGCAAAAGAA.....CACGTGATGCTCGCGC 3118
1146 yAspAlaSerPheSerLeuGlyAsnThrGlyLysPheValAspLeuGly 1163
3119 CGTGGCGTTACCAACTCATCCGCAAGACGGC.....GAGTTCCGCTG 3162
1163 hrTyrGluTyr...ValLeuLysSerAspGlyAsnSerAsnThrPheLeu 1178
3163 CATAAATCCGTCAAGAAAGAGCTTTCGACAAACTCGGCAAGGCAGA 3212
1179 ThrAsnAspValIysPro...AsnProAspProAsnProAsnProAs 1195
3213 AGCCAAAACAGAGCGGAAAAAGACAAACGCGAAAGCTTGACGGGCTCA 3262
1195 snProLys..... 1197
3263 TTGCGCGCGGCGCGATGCGTGCAGAAAAGACAGAAAGCTTGCAGACG 3312
1198 .....ProAsp... 1199
3313 GCCCGCAGCAGAGCGGGGAAATGTCCGATTATGACAGCGGAGGAAGA 3362
1199 ..... 1199
3363 GAAAAACGGGTGACAGCGGATTAAGACACGCTTGGGAAACAGCGCG 3412
1199 ..... 1199
3413 AAGCGAAACCGCGCGCTACACGCTTCCCGCGCGCGCGCGCGC 3462
```

```
1200 .....ProLysProAsp...ProLysProAspProL 1209
3463 CGCGGGGATTGGCCGAACGTCAACCCCAACCCGACCCGACGAGCG 3512
1209 ysrPro...Asp...ProThrProGluProThrProThrProValProGlu 1224
3513 CGACCTGATCAGCGGCTTATGCCAATAGCGGTTTGAATTTTCGCCCA 3562
1224 sarGluLeuPro.....SerThrAlaValLeuAsnMetAlaIat 1239
3563 CGCTCAACAGCGTTTC.....GCCGTACAGACCAATTA 3597
1239 hrLeuProLeuValPheAspAlaGluLeuAsnSerIleArgGluArgLeu 1255
3598 GACCGCGTATTTCGCAAGACCGCGCAACGCGGCTTTGGACAAAGCGCAT 3647
1256 AsnIleMetLysAlaSerProHisAsnAsnValTyr...GlyAlaThr 1271
3648 CCGGGACACCAAAACACTACCGTTCCGACAGATTTCGCGCTACCGCCAC 3697
1271 rTyrAsnThrArgAsnAsnValThrThrAspAlaGlyLysPheGluG 1288
3698 AAACGACCTGCGCCAAATCGSTATGCAGAAAACCTC..... 3735
1288 LThr.....LeuThrGlyMetThrValGlyIleAspSerProAsn 1301
3736 .....GCAAGCGGCGCGCTCGGCATCTGTTTTCGCAACACG 3773
1302 AspIleProGluGlyIleLeuThrLeuGlyAlaPheMetGlyTyrSerH 1318
3774 GACGAAAAACACCTTCGACGACGCG..... 3798
1318 serHisIleGlyPheAspArgGlyGlyHisGlySerValGlySerTyrS 1335
3799 ..ATCGCAACTCGGACGCGCTTCCGACGCGCGCTTTC..... 3837
1335 erLeuGlyGlyTyrAlaSerTrpGluHisGlySerGlyPheTyrLeuAsp 1351
3838 GGGCAATACGGCATGACAGGTC.....TACATCGCATACAGCGC 3878
1352 GlyValValLysLeuAsnArgPheGluSerAsnValAlaGlyLysMetSe 1368
3879 GGGCGCGGGTTTTCGACGCGCAGCCTT...TCAGACGCGATCGAGGCA 3925
1368 rSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnLysLeuGlyLys 1385
3926 AAATCCGCGCGCGCGCTGCTGCATTACGGCATTCAGGACGATACCGGCGC 3975
1385 ls.....IleGluThrGlyMetArgPhe 1392
3976 GGTTCGCGGATTCGGCATCGAACCGGACATCGCGGACGCGCTATT 4025
1393 ThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuThrGlyPheTh 1409
4026 CGTCCAAAACGCGGATTACCGC.....TACGAAAACGCTCATA 4063
1409 rAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGluSerLysServ 1426
4064 TCGCCACCGCGCGGCTGATTCACACCGCTACCGC...GGGCGCATTAAG 4110
1426 alaThrThrArgSerIle.....TyrArgGluLeuGlyAlaThr 1438
4111 GCAGATTATTCATTCAAAACGCGCGCAACGCGATTCGATCCAGCGCTATT 4160
1439 LeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGluProTyrPle 1455
4161 GAGCTGTCTCATACGATGCGCGCTTGGCAAAAGTCCGAACAGCGCTCA 4210
1455 ulysAlaIleValArgLysGluPheValAspAspAsnArgValLysVal 1472
4211 ATACC..... 4215
1472 snAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArgGlyIleTyr 1488
```

```

4216 ...GCCGATATGGCTCAGATTTGCGCAAAACCGCAGTCGGAAATGGG 4262
      |||  :::::|||||  |||  :::::  ||
1489 GlnIaGlyIleYsaIaSerPheSerThrIleuSerGlyHisLeuGI 1505
      |||  :::::|||||  |||  :::::  ||
4263 CGTAAAGCCGCAATCAAGGTTTCACGCTGTCCTCCAGCGTCGCCCG 4312
      |||  :::::|||||  |||  :::::  ||
1505 yVal.....GlyTyrSer.....HisGlyAlaGly 1514

4313 CCAAGGCCCG 4323
      :::::|||||
1514 alGIuserPro 1517

seq_name: pir2:A65044

seq_documentation_block:
hypoetical protein b2647 - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #extL_change 08-Oct-1999
C/Accession: A65044
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617
A/Accession: A65044
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1569 <BLAST>
A/Cross-references: GB:AE000350; GB:U00096; NID:q2367147; PIDN:AAC75695.1; PID:q1789001;
A/Experimental source: strain K-12, substrain MG1655

alignment_scores:
Quality: 364.50 Length: 1634
Ratio: 0.485 Gaps: 81
Percent Similarity: 46.022 Percent Identity: 19.951

alignment_block:
US-09-303-518D-649 x A65044 ..

Align seg 1/1 to: A65044 from: 1 to: 1569

445 GGGAGTAAGGCCATCCTATGCGCGCATATCATATGCGCGCTTGC 494
      |||||  :::::  :::::  :::::  :::::  ||
97 GlyThrThrAsnAsnThrHisIleLeuHisGlyGlnGlnAsnValHi 113
      |||  :::::|||||  :::::|||||  :::::|||||
495 TAAATTTGTACAGATGCGAAGCACTGTGAAATGACCAAGTATATGAT 543
      |||  :::::|||||  :::::|||||  :::::|||||
113 sGIyGlyValSerAsnGlySerLeuIleGIuSerGlyGlyTyrGlnAspI 130
      |||  :::::|||||  :::::|||||  :::::|||||
544 .....GGCGGAAATATATCATGCAAAATTAATACCGTCGCGTGT 585
      :::::|||||  :::::|||||  :::::|||||
130 IeGIySerHisAsnAsnPhenValGIyGlnAlaAsn.....AsnThr 143
      |||  :::::|||||  :::::|||||  :::::|||||
586 CGATTTGGGCGAGCGACGCAATATGCGCATCTGATGAAGATGAGCCCA 635
      |||  :::::|||||  :::::|||||  :::::|||||
144 ThrIleAsnGlyGlyArgGln..... 150
      |||  :::::|||||  :::::|||||  :::::|||||
636 TAAACCGGAAGTTTCATATCATATTGCA.....AGTGGTATTTCTGGC 679
      |||  |||  |||  :::::  |||  :::::  ||
151 .....SerIleHisAspGIyGIyIleSerThrGlyThrThr 163
      |||  |||  |||  :::::  |||  :::::  ||
680 TCGTTGGGCAATACCTTTCGACAAATGCA.....TCAGGTGGTGGC 723
      :::::|||||  :::::|||||  :::::|||||  :::::|||||
163 IeGIuSerGlyAsnGlnAspValTyrGlyGlyGlyIleSerAsnGlyThr 179
      |||  :::::|||||  :::::|||||  :::::|||||
724 ACAGTCACTAGTAGTGAATAAATTA.....CATAGCCCATATGCTTT 770
      |||  :::::|||||  :::::|||||  :::::|||||
180 ThrIleGlyGlyIleAlaSerArgValGIuGlyGlySerAlaAsnGIyI 196
      |||  :::::|||||  :::::|||||  :::::|||||
771 TTTCACAACAGAGGCTCA..... 789
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

196 eleuIleAspGIyGlySerGlnIleValIyValGIuGlnIyHisAlaSpG 213
790 .....TTTGGCAGACGTGGCTCCACCAATGTTATGATGATGCCCA 831
      |||  :::::|||||  |||  :::::  ||
213 LyThrThrIleAsnIySerGlySerGlnAspValIaGIuGlySerIeu 229
      |||  :::::|||||  |||  :::::  ||
832 AAGCAAAAGTGGTAAATTAATGCGGTATTCGCAAGGCCGACCCCTATAT 881
      |||  :::::|||||  |||  :::::  ||
230 AlaThrAsnThrThrIleAsnGly.....GlyArgGlnTyrVa 242
      |||  :::::|||||  |||  :::::  ||
882 AGCA.....AAAGCATGGCTTCGACG 904
      |||  :::::|||||  |||  :::::  ||
242 IGIuGlnSerThrValGIuThrThrThrIleIyAsnGlyGIuGlnIa 259
      |||  :::::|||||  |||  :::::  ||
905 TGGTTGCTAAAGATGGTCTCTATGATGAATCTTGTCTGAGATACCAT 954
      |||  :::::|||||  |||  :::::  ||
259 rGlyAllyGIuSerArgAlaIleAspThrThrIleGIuGIyGlyThrGln 275
      |||  :::::|||||  |||  :::::  ||
955 TCAGTATTTACGACGACGCAAGTCAAAATGGGAAATCTTTTACGAC.. 1002
      |||  :::::|||||  |||  :::::  ||
276 SerIleuAsnSerIySerThrAlaIyAsnThrHisIleTyrSerGIyGI 292
      |||  :::::|||||  |||  :::::  ||
1003 .....GATTAATATGGCGACAGCAAAATTCATGGCCAAACATG 1039
      |||  :::::|||||  |||  :::::  ||
292 yThrGlnIleValAlaAspAsnThrSerThrSerAspValIleGIuValTyrS 309
      |||  :::::|||||  |||  :::::  ||
1040 AACACAATCTCTGCTCTAATAGA..... 1062
      |||  :::::|||||  |||  :::::  ||
309 eGIyGlyValIleuAspValArgGIyGlyThrAlaThrAsnValThrGln 325
      |||  :::::|||||  |||  :::::  ||
1063 .....TTAAACACGACCCGTTCAATGTTTAAGTTTC 1097
      |||  :::::|||||  |||  :::::  ||
326 HisAspGIyAlaIleuIySerThrAsnThr.....AsnGIyThrThr 339
      |||  :::::|||||  |||  :::::  ||
1098 TTTATCCGAGACAGCAAGAACCTGTTAT.....CATGCTG 1135
      |||  :::::|||||  |||  :::::  ||
339 rValSerGIyThrAsnSerGIuGIyAlaPheSerIleHisAsnHisValA 356
      |||  :::::|||||  |||  :::::  ||
1136 CA.....GCTGGT.....GTCAACAGTTAT 1155
      |||  :::::|||||  |||  :::::  ||
356 IAspAsnValIleuLeuGIuAsnGIyGlyHisIleuAspIleAsnAlaTyr 372
      |||  :::::|||||  |||  :::::  ||
1155 ..... 1155

373 GlySerAlaAsnIyThrIleIleIyAspGIyThrIleSerValIle 389
      |||  :::::|||||  |||  :::::  ||
1156 .....CGACCCAGACTGAATATAGA..... 1176
      |||  :::::|||||  |||  :::::  ||
389 uThrAsnAlaIyAlaAspAlaThrArgIleAspAsnGIyGlyValIle 406
      |||  :::::|||||  |||  :::::  ||
1177 .....GAAAT 1182
      |||  :::::|||||  |||  :::::  ||
406 sPValAlaGIyAsnAlaThrAsnThrIleIleAsnGIyGlyThrGlnAsn 422
      |||  :::::|||||  |||  :::::  ||
1183 ATTTCCTTTATTCGCAAGGAAAGCGCAATGTACTTACCAACAAT 1232
      |||  :::::|||||  |||  :::::  ||
423 IleAsnAsnTyrGIyIleAlaThrGIy.....ThrAsnI 424
      |||  :::::|||||  |||  :::::  ||
1233 CAATCAAGT.....GCTGAGCATTAATTTCCACAGAG 1267
      |||  :::::|||||  |||  :::::  ||
434 eAsnSerGIyThrGlnAsnIleIySerGIyGly.....LysAla 448
      |||  :::::|||||  |||  :::::  ||
1268 ATTTTACGGTCCCTCGAATAATAC.....GAAACTTG 1302
      |||  :::::|||||  |||  :::::  ||
448 sPThrThrIleIleSerSerGIySerArgIleValIaGIuGIyAspGIy 464
      |||  :::::|||||  |||  :::::  ||
1303 CAGGCGGCGGCTTCATATCATGATGAAGACAGTACCTTACTTGGAAAT 1352
      |||  :::::|||||  |||  :::::  ||
465 ThrAlaIleGIySerAsnIleSerAlaGIyGlySerIleuIleValTyrTh 481
      |||  :::::|||||  |||  :::::  ||
1353 AACGGCGTGGCAACGAC.....CGCTGT 1378
      |||  :::::|||||  |||  :::::  ||
481 rGIyGIyIleAlaHisGIyValaAsnGlnGIuThrGIySerAlaIleuValA 498
      |||  :::::|||||  |||  :::::  ||

```

```

1379 CCAAAATCGCAAGGCACG.....CTGCAC 1404
      ::::: ||| |||||
498 laasnthrglyalaglyThrAspIleuglyTyraSnlyLeuSerHis 514
      ::::: ||| |||||
1405 GTTAAAGCCAAAGG.....GAAACCAAGG 1430
      |||
515 PheThrIleThrIleGlyGlyGlyAlaAsnTyValValIleuGlnAsnThrG 531
      ::::: ||| |||||
1431 CTGCATGAGGTGGCGGACGCTACGTATTTGGATCGACGAGCAGC 1480
      ::::: ||| |||||
531 yGluLeuThrValValAlaLysThrSerAlaLysAsnThrThrIleAsp 548
      ::::: ||| |||||
1481 ATAAAGGCAAA.....AAACAAGCCTTTAGTCAAAATCGGCTTG 1518
      |||||
548 hrGlyGlyLysLeuIleValIleGlnLysGlnAlaLysThrAspSerThrArg 564
      ::::: ||| |||||
1519 GTCAAGCGGCGGCTACGCTGCTGCACTGAATGCCGAT..... 1554
      ::::: ||| |||||
565 LeuAsnAsnGlyGlyValLeuGlnValGlnAspGlyGlyAlaLysSH 581
      ::::: ||| |||||
1555 ....AATCAGTTCAACCCGACAACTCTATTTCGCTTCGGCGGAC 1600
      ::::: ||| |||||
581 sValGlnGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSerGlyT 598
      ::::: ||| |||||
1601 GTTTC.....GATTTAAAGGGCATTCGCTTCGCTCCACCGTATTCAA 1644
      |||
598 hrLeuIleGlnGlyThrAsnSerTyGlyAspAlaPheTy.....IleArg 613
      ::::: ||| |||||
1645 AATACGATGAAGGGCGATGATGTCACCAACCAATCAAGACAAAGAATC 1694
      |||||
614 AsnSerGlnAlaLysAsnValValLeu.....GluAsnAlaGlySe 627
      ::::: ||| |||||
1695 CACCGTTACCATTTACGCAATTAAGATTTGCTACACCGGCAATATACA 1744
      ::::: ||| |||||
627 rLeuThrValValThrGlySerArgAlaValAspThrIleIleAsnAla 644
      ::::: ||| |||||
1745 ACAGCTGTGATGCAAAAGAAATGGCTACACGCTTGCTGGTCGAG 1794
      |||
644 sn.....GlyLysMetAspVal.....TyGly... 651
      ::::: ||| |||||
1795 AAAGAT.....ACGACCAAAAGAGCGGCGCTCAACCTTGTTTACCA 1838
      |||||
652 LysAspValGlyThrValLeuAsnSerIleGlyThrGlnThrIleTyAl 668
      ::::: ||| |||||
1839 GCGCGCGGAGAAAGACCGCAGCTCTGCTTCGCGGCAACAAATTTAA 1888
      ::::: ||| |||||
668 aserAlaThrSerAspLysAlaAsnIleLysGlyGlyLysGlnThrVal 685
      ::::: ||| |||||
1889 ACGGCAACATCACGCCAAACAAC.....GGCAAACTGTTTTCAGC 1929
      |||
685 yGlyLeuAlaThrGlnAlaAsnIleGlnSerGlyGlnGlnIleValAsp 701
      ::::: ||| |||||
1930 GGCACACCAACACCGCAGCCTTACATTTAAACGAC..... 1968
      |||
702 GlyGlySerThrGlnLys.....ThrHisIleAsnGlyGlyThrGln 716
      ::::: ||| |||||
1969 .....C 1969
      ::::: ||| |||||
716 rValGlnAsnTyGlyLysAlaIleAsnThrAspIleValSerGlyLeuG 733
      ::::: ||| |||||
1970 ATTGGTCGCAAAAGAGGCGATTCCTCGCGGGAATCGTG...TGGGAC 2016
      ::::: ||| |||||
733 GlnIleIleMetAlaAsnGlyThrAlaGlnGlySerIleIleAsnGlyGly 749
      ::::: ||| |||||
2017 AACGAGCTGATCAACCGCACATTTAAAGCGAAACTTCGAAATTTAAAG 2066
      ::::: ||| |||||
750 SerGlnValValAsnGlnGlyGlyLeuAlaGlnAsnSerValLeuAsn 766
      ::::: ||| |||||
2067 CGGACAGCGCGGTGTTCCGCAATTTGCGAAAGTCAAGGCGATGGC 2116
      |||
766 pGlyGlyThrLeuAspValArgGlyLysGlySerAlaThrGly..... 780
      ::::: ||| |||||

```

```

2117 ATTGAGCAATCAAGCCCAAGCAGTTTGTGTGCGACCGCATCAAGC 2166
      ::::: ||| |||||
781 ..IleGlnGlnSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThr 796
      ::::: ||| |||||
2167 CACGCAATCTGTACAGCTTCGCTGCGACTGACGGCTGCACAAATGTGCGA 2216
      ::::: ||| |||||
797 ArgValThrGlyThrArgAlaAsp.....GlyValAlaPheSerIleG 811
      ::::: ||| |||||
2217 AAAA.....ACCA 2224
      ::::: ||| |||||
811 uGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyValLeuThrV 828
      ::::: ||| |||||
2225 TTACCGACAT..... 2235
      ::::: ||| |||||
828 AlGlnSerAspThrSerAspLysThrGlnValAsnMetGlyGlyArg 844
      ::::: ||| |||||
2236 .....AAAGTATGCTTCATGATAGTAAGCCGACATCAGCGCA 2276
      |||||
845 GlnIleValLysThrLysAlaThrAlaThrGlyThrThrLeuThrGlyG 861
      ::::: ||| |||||
2277 TGTGATCTT.....GCCGATCCGCTCATTTAAATCTCAGAC 2314
      ::::: ||| |||||
861 yGlnGlnIleValIleGlnGlyValAlaAsnGlnThrThrIleAsnAspGlyG 878
      ::::: ||| |||||
2315 GGCTTGCCACACTCAACGCGCATCTTAGTGCAAATGGCGAT..... 2355
      |||||
878 LysIleGlnThr.....ValSerAlaAsnGlyGlnAlaIleLys 890
      ::::: ||| |||||
2356 ACAGCTTATACAGTACGACCAACGCCACCCCAAGCGCAAC..... 2397
      ::::: ||| |||||
891 ThrLysIleAsnGlnGlyGlyThrLeuThrValAsnAspAsnGlyLysAl 907
      ::::: ||| |||||
2398 ..CTTACCTCTGTGGCGATGCCCAAGACAACTTAATCAACGCAATTA 2446
      ::::: ||| |||||
907 arThrAspIleValIleAsnSerGlyAlaAlaLeuGlnThrSerThrAla 924
      ::::: ||| |||||
2447 ACGGCAACACATCGGCTTCGGCGCAT.....GCTTCATTTAACTA 2487
      |||||
924 snGly...IleGlnIleSerGlyThrHisGlnTyGlyThrPheSerIle 939
      ::::: ||| |||||
2488 ACGGACCGCGCGTCAAAAGCGGAGTCTGACGCTTTCGCGCAACGCTAA 2537
      |||
940 SerGlyAsnLeuAlaThrAsnMetLeuLeuGlnGlnGlyAsnLeuLe 956
      ::::: ||| |||||
2538 GCGCAACGTAAGCATTCGCGCATCAAGCTCAAGTAACTCTCCCTACCGCAT 2587
      ::::: ||| |||||
956 uValLeuAlaGlyThrGlnAlaArgAspSerThrValGlyLysGlyGly 972
      ::::: ||| |||||
2588 ACGCATATTCATTTTGAAGCAGCGCTTACCGGCAAAATCAGCGGC 2637
      |||||
973 ..AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsnSerGly 988
      ::::: ||| |||||
2638 GGCAGGATATCGGCTTACACTTTAAAGACAGCAATGACGCTGCCGTC 2687
      ::::: ||| |||||
988 GlyGlnTyThrLeuGlyArgSerLysAspLysPheGlnAlaLeuAla 1005
      ::::: ||| |||||
2688 AGGACGGAATTA..... 2700
      ::::: ||| |||||
1005 gAlaGlnAspLeuGlnValAlaGlyGlyThrAlaIleValTyAlaGlyT 1022
      ::::: ||| |||||
2701 .....GGCAATTTAACTT... 2715
      ::::: ||| |||||
1022 hrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSerLeuMet 1038
      ::::: ||| |||||
2716 .....GACACGGCACCC...ATTACACTCAATTCGCGCTATGCG... 2751
      ::::: ||| |||||
1039 ThrProArgAspAsnValThrProValLysLeuGlnGlyAlaValArgI 1055
      ::::: ||| |||||
2752 ..CAGATTCGCGGCGGCGGCAACCGGCACTGCGACAGATGCGCGCGCG 2800
      ::::: ||| |||||
1055 eThrAspSerAlaThrLeuThrLeuGlnGlyAsnGlyValAspThrThrLeu 1072
      ::::: ||| |||||
2801 GCCGTTGCGCGCGTTCGCGCGGT..... 2823
      ::::: ||| |||||

```

```

1072 laaspleuthrAlaAlaSerArgIySerValItrPleuAsnSerAsn 1088
      :::: |||
2824 .....TCCCTATTA.. 2832
      |||||
1089 SerCysAlaGlyThrSerAsnCysGIuTyrArgValAsnSerLeuLeu 1105
      |||||
2833 .....TCCGTTACACCGCCCACTTCGG 2854
      ::::: |||
1105 uAsnAspGIyAspValItyrLeuSerAlaGlnThrAlaAlaProAlaThrT 1122
      ::::: |||
2855 TAGAATCCCGTTCAACAGCGTAAAGCGCAATTAAGAGGTCAG 2904
      ::::: |||
1122 hTrasnGlyIleTyrAsnThrLeuThrThrasn...GIuLeuSerGIyS 1137
      ::::: |||
2905 GGAACATTCGCTTATGTGGAACCTTCGCTACCGCAGCACAAT 2954
      |||:: |||
1138 GIyAsnPhetYrLeuHsthrAsnValAlaGlySerArgGIyAspGln 1154
      ::::: |||
2955 GAAAGCTGGCGAAAGTTCCGAAGGCACTTACACCTGGCGTCAACA 3004
      |||:: |||
1154 uValValAsnAsnAlaThrGIyAsnPhetYsIlePhetValGlnspt 1171
      ::::: |||
3005 CCGGCAACGAACCTCGAAGCCTCGAACATTTGAGTAGGAGAA 3054
      |||:: |||
1171 hTrGIyValSerProGlnSerAspAspAlaMetThrLeuVal..... 1184
      ::::: |||
3055 GACAACAACCGCTGTCCGAAACCTTAATTTACCTCGCAAAAGCA 3102
      |||:: |||
1185 .....LysThrGIyGIyGIyAspAlaSerPhetThrLeuGIyAsnThrGI 1199
      ::::: |||
3103 ....CAGGTGATGCGGCGGCGGTTCACATCATCGCAAAAGACG 3148
      ||||| |||||
1199 yGIyPhetAlaAspLeuGIyThrTyrGIyTyr...ValLeuIySerSpg 1215
      ::::: |||
3149 GCGAGTTCGCGCTCATTAATCCGTCAAAGAACAGCTTTCCGACAAA 3198
      |||:: |||
1215 Ly.AsnSer..... 1217
      ::::: |||
3199 CTGCGCAAGCAGAACCAAAACAGCGGAAACAGAACAGCGCAAG 3248
      ::::: |||
1218 .....AsnTrPA 1220
      ::::: |||
3249 CCTTGACGCGTGATTCGGCGCGCGGATCCGTCGAAAGACAGAAA 3298
      ||||| |||||
1220 snLeuThrAsnAspValIySProAsnProAspPro..... 1231
      ::::: |||
3299 GCGTTGCCGAACCGGCGCGGACGAGCGGGGAAATGTGCGCATATG 3348
      ::::: |||
1232 ...IleProAsn..... 1234
      ::::: |||
3349 CAGCGGAGGAGAAAGAAAAAGGGTGACGGGATAAAGACACCGCTT 3398
      ::::: |||
1234 ..... 1234
      ::::: |||
3399 GCGGAACAGCGCGAAGCGGAAACCGCGGCTACACCGCTTCCCC 3448
      ||||| |||||
1235 .....ProIySProAspProIySProAspProIySPro 1245
      ::::: |||
3449 GGGCGCGCGCGCGCGCGGATTCGGCAACTGCACACCCCAACCGAG 3498
      |||:: |||
1246 AspPro...AsnProIySProAsp...ProThrProAspProThrProThr 1260
      ::::: |||
1261 ProValProGlnIyArgIleThrPro.....SerThrAlaAlaValIle 1275
      ::::: |||
3549 TGAATTTTCCGCGCAGCTCAACAGCTTTC.....GCGC 3583
      ::::: |||
1275 uAsnMetAlaAlaThrLeuProLeuValPhetAspAlaGlnLeuAsnSerI 1292
      ::::: |||
3584 TACAGAGCAATTAGACGCGGTATTTGCCGAGACCGCGCAACCGCTT 3633
      ::::: |||

```

```

1292 IeArgGIuArgLeuAsnIleMetLysAlaSerProHIsAsnAsnVal 1308
      ::::: |||
3634 TGGACACCGGCATCCGGGACACCAACACTACCGTTCCGAACTTCCG 3683
      |||:: |||
1309 Trp...GIyAlaThrTyrAsnThrArgAsnAsnValThrThrAspAlaGI 1324
      ::::: |||
3684 GCGCTACCGCCCAACAAACCGACCTGCGCCCAATGCGTATCGCAAAAAC 3733
      |||:: |||
1324 yAlaGlyPhetGIyGlnThr.....LeuThrGIyMetThrValGIyI 1338
      ::::: |||
3734 TC.....GCGACGCGCGCGCTGCGCATCTG 3759
      ::::: |||
1338 IeAspSerArgAsnAspIleProGlnGIyIleThrThrLeuGIyAlaPh 1354
      ::::: |||
3760 TTTTGGCAACACCGGCAAAACACCTTCGACGACGCG..... 3798
      ::::: |||
1355 MetGIyTyrSerHIsThrIleGIyPhetAspArgGIyGIyHIsGIyS 1371
      ::::: |||
3799 .....ATCGGCACCTCGCACAGCGCTTCCACGCGCGCG 3832
      ::::: |||
1371 rValGIySerTyrSerLeuGIyTyrAlaSerTrpGIyHIsGIySerG 1388
      ::::: |||
3833 TTTTC.....GGCAATACGCGCATCGACAGTTTC.....TAC 3864
      |||:: |||
1388 IyPhetYrLeuAspGIyValValIyLysLeuAsnArgPhetLysSerAsnVal 1404
      ::::: |||
3865 ATCGGCATCGACGCGCGCGGCTTTAGCAGCGGACGCTT...TCAGA 3911
      |||:: |||
1405 AlagIyLysMetSerSerGIyGIyAlaAlaAsnGIySerTyrHIsSerS 1421
      ::::: |||
3912 CCGCATCGGACCAAAATCCGCGCGCGCTGCTGCATTAACGCAATTCAG 3961
      ::::: |||
1421 nGIyLeuGIyGIyHIs.....IleGIyUT 1429
      ::::: |||
3962 CAGCATACGCGCGCGGTTGCGCGATTCGCGATCGCAACCGCACATTCG 4011
      ::::: |||
1429 hTrGIyMetArgPhetThrAspGIyAsnTrpAsnLeuThrProIyTrAlaSer 1445
      ::::: |||
4012 GCACGCGCTATTTGTCGCAAAAGCGATTAACCGCGCAACATTCGC 4056
      |||:: |||
1446 LeuThrGIyPhetThrAlaAspAsnProGIyTyrHIsLeuSerAsnGIyMe 1462
      ::::: |||
4057 .....GTCAATACGCGCACCGCGCTTCGATTAACCGCTACCGC. 4098
      ::::: |||
4099 ...GCGGCATTTAAGCGCATTAATTCATAACCGCGCAACATTCGC 4146
      |||:: |||
1462 LysSerLysSerValAspThrArgSerIle.....TyrArgG 1475
      ::::: |||
1475 IuLeuGIyAlaThrLeuSerTyrAsnMetArgLeuGIyAsnGIyMetGIu 1491
      ::::: |||
4147 ATCAGCGCTTATTTAGCGCTGTGCTTACCGATGCGGCTTCGGGCAAA 4196
      ::::: |||
1492 ValGIuProTrpLeuLysAlaAlaValArgLysGIuPhetValAspAsp 1508
      ::::: |||
4197 CCGACACGCGCATATACGCGCTATTTGGCTCAGATTC.....GGCA 4240
      ||||| |||||
1508 nArgValIyValAsnSerAspGIyAsnPhetValAsnTyrLeuSerGIyA 1525
      ::::: |||
4241 AAACCGCAGTGGCAATGGCGGCTAAACCGCAAAATCAAGGTTTCAG 4290
      ::::: |||
1525 rGIyArgIyIleTyrGlnAlaGIyIleLysAlaSerPhetSer...Thr 1540
      ::::: |||
4291 CTGTCCTCGACGCTGCGCGCGCAAGCGCCCACTGAGAACGCAACA 4340
      ||||| |||||
1541 LeuSerGIyHIsLeuGIyValGIy.....TyrSerH 1551
      ::::: |||
4341 CAGCGCGCGCATCAAA.....TTAGCTACCGCTG 4371
      ||||| |||||
1551 sSerAlaGIyValGIuSerProTrpAsnAlaValAlaGIyValAsnTrp 1567
      ::::: |||

```

seq_name: p1r2:B90674
seq_documentation_block:

Aldh-I adhesin-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90674
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gatawara, N.; Ysunaga, T.; Kihara, S.; Shiba, T.; Hatfort, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Accession: B90674
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1327 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033785.1; PID:G13359819; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS0362

alignment_scores:

Quality:	325.00	Length:	1474
Ratio:	0.509	Gaps:	72
Percent Similarity:	43.284	Percent Identity:	20.692

alignment_block:

US-09-303-518D-649 x B90674 ..

Align seg 1/1 to: B90674 from: 1 to: 1327

```
712 TCAGTGTGGCAGACGTCACCTAGTAAATTAACATAGCCC 761
|||||
96 SerGlyGlyGlyLeuAspIleThrGlyValThrIleAsnGlyPr 112
762 ATATGTTTATACCAACAGAGGCTCATTT.....GGCGACATG 802
|||||
112 OLeuAsnPhleuLeuLysGlyThrGlyPheLeuAsnValSerSnaIag 129
803 GGTCCACCATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAT 852
|||||
129 LYSerGlyLeuThrAlaAspAspLeuThrGlySer.....Asn 141
853 GGGGATTTGCAACGCGCACCCCTATATAGAAAAAGCATGGCTTCCA 902
|||||
142 SerGlyMetArgHisAspArgGlyThrPheAsnValSerAsnGlyGly 158
903 GCTG..... 906
158 sIleHisValLysGlyThrSerArgLeuThrTyrlLeuGlnGlyAsnValS 175
907 ..... GTTGTAAAGTTGGTCTATGATGAATCTTT 939
175 eGlyGlnGlySerGlnValAsnSerGlnThrPhePheMetGlyValTyr 191
940 GCTGGAGATACCATTCAGTATCTACGACCAACGTCAAATGGGAATA 989
192 .....GlySerTy 194
990 CTCCTTTTAC.....GACGATATATGCGACAGAAATTCATG 1030
|||||
194 rGlyGlyAsnGlnTyrLeuSerValAsnAsnGly...GlyGlnValAsnA 210
1031 CCAACATGACACACAT..... 1047
210 laArgLysGlnIleSerLeuGlyTyrTyrAspGlnValSerAspThrThr 226
1048 .....TCTGTCCCTATAGATTAAACACAGAACCTGTAATG.. 1086
227 LeuAlaValSerGlnGlyGlyIleSerAlaProThrIleSerLeuSe 243
1087 ..TTAATGTTCTTATCCGAGACAGAACAGAACTGTTATTCAGCTG 1135
|||||
243 rThrAsnSerGlnLeuAlaLeuGlnGlySerAlaAlaLysA 260
```

```
1136 CAGGTGCTGTCACACGTATTCAGCCGACGTAATATGAGAAATATT 1185
|||||
260 laAlaGlyIle.....IleAspIleGlyIle 269
1186 TCCCTTATTCAGCAGAGAAAGGCAATGATCTATTCACAGAAC..... 1230
270 GluPheValThrAlaLysThrSerGlyLysIleThrLeuAsnHisTh 286
1230 ..... 1230
286 rAspLysAspAlaThrIleSerAlaAspIleValSerGlySerGlyL 303
1231 .....ATCAATCAAGTCTGGAGATTTATTTCCAGAGATTTT 1272
303 euGlyTyrIleAsnAlaLeuAsnGlyThrThrTyrlLeuThrGlyAsp... 318
1273 ACGTCTCGCTGCAAAATACGAACTTGCAAGCGCGGCGTTCATAT 1322
319 .....AsnSerAlaPheSerGlyLys...ValLysI 328
1323 CAGTGAAGACAGTACCGTT.....ACTTGAAG 1351
328 eGlnGlnAsnGlyAlaLeuGlyIleThrGlnAsnIleGlyThrAlaGlu 345
1352 TAAACGGCGTGGCAAAAGCAGCCGCTGTCCAAATGCGCAAGGACGCTG 1401
345 IeAsn.....AsnArgGlyLysLeu 351
1402 CAGCTTCAAGCC.....AAGGGGAAACCA 1427
352 HisLeuLysAlaAspAspSerMetThrPheAlaAsnLysIleSerGlyAs 368
1428 AGGCTGATCAGCTGGGCGAGCGTACGATCTTTGGATGACAGCGCAG 1477
368 nGlyThrIleSerIleAspSerGlyThrValGlnLeu..... 380
1478 ACGATAAAGCAAAACACAGCTTACT...GAATCGCTGGTGCAGC 1524
381 .....ThrGlyAsnAsnTyrAlaPheSerGlyTyrIleAspValAlaSer 395
1525 GCGCAG.....GTAAGGTGCACACT 1544
396 GlyAlaValAlaValIleSerGluAspLysAsnIleGlyAlaGlnIle 412
1545 GAATGCCGATATCAGTTC.....ACCCGACAAATCTATT 1582
412 uAspValAspGlyLysLeuGlnIleAsnAlaAsnLysAspThrValPheA 429
1583 TCGGCTTTCGGCGCGAGCT.....TTGATTTTAAACGGGCAAT 1620
429 sPAsnAspLeuGlnGlyArgGlyIleValGlnIleAsnMetGlyAsnHis 445
1621 TCGCTTGTTCACCGCTATTCAAATACGAT.....GAAGGGCGAT 1664
446 GluPheSerPheAspGluPheAlaTyrThrAspThrPheGlnGlySerI 462
1665 GATTGTCAACCAATCAAGACAAAGAAATCCACGTTACCATTCACAGCA 1714
462 uAlaPheGlnAsn..... 466
1715 ATAAAGATATTGTCACACCGGCAATTAACAACAGCTTGATACCAAAAA 1764
467 .....ThrThrPheAsn.....LeuGlnLysAsnAla 475
1765 GAATTTGCTACACGCTTGTTGGCGAGAAAGATACCAAAAGCAA 1814
476 GluPheLeuGlnLysGly.....GlyIleThrAlaGlyL 487
1815 CGGCGGCTCAACCTGTTTACCAAGCCCGCGAGAGAACCGGACCCGCG 1864
487 nGlySerLeuValThrValGlyGlyAlaHisSerIleSerThrLeuG 504
1865 TGCCTTCCGGGGGAACAATTTAAACGGCAACATCAG..... 1902
```



```

3412 GAAGCGGAAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGC 3461
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
969 GUPProsprometProaspnProgluProaspnProasn..... 981
3462 CCGCGCGGATTTCCCGCACTGACACCGCACCGCGCGCGCGCGC 3511
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
982 .....ProgluProaspnProaspnProthProthProgluP 994
3512 GCGACCTG.....ATCAGCGCTTATGCCAATAGCGGTTTG 3546
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
994 rospneuspasnValaspasnpleuArgprogluAclYserTyrIle 1010
3547 AGTGAATTTTCCGCGCACGCTCAACGCTTTTCCG.....GTCAGGA 3590
1011 AlasnleuAlaAlaAla...AsnThrMetPheThrThrArgLeuHisGI 1026
3591 CGAATTA.....GACCGCGTATTTCGCGAAGACCGCC 3622
1026 uArgLeuGlyAsnThrTyrTyrThrAspMetValThrGlyGlnLysG 1043
3623 GCAACGCGCTTTGGACACAGCGGATCCGGGACACCAACACACGCTTCG 3672
1043 LnthrThrMetTrpMetArg.....HisGlyGlyHisAsnLysTrp 1057
3673 CAAGATTTCCGCGCGCTACCGCACAAACCGACCTGCCCAATCGTAT 3722
1058 ArgAsp..... 1059
3723 GCGAATAAACCTCGGACGCGCGCGCTGGCATCTGTTTCCGACAAAC 3772
1060 .....GlySerGlyGlnLeu.....LysT 1066
3773 GGAACGAAACACCTTCGACGCGCATCGCACTCGCACGCGCTTGGC 3822
1066 hArgLeuAsnArgTyrValLeuGlnLeuGlyAspVal..... 1079
3823 CACGCGCGCGTTTCCGGCATACGCGCATCGACAGTTTTCATCGGCAT 3872
1080 .....AlaGlnTyrSerGlnAsnGlySerAspArgTrpHisValGlyVa 1094
3873 CACGCGCGCGCGCGGTTTACG...ACGCGGACGCTTCAGACGCGCATCG 3919
1094 LMetAlaGlyTyrGlyAsnSerAspSerLysThrIleSerSerArgThrG 1111
3920 GAGGCAAAATCCGCGCGCGCTGCTGCATTAC.....GGCATTCAGGCA 3963
1111 LTTTArgAlaLysAlaSerValAsnGlyTyrSerThrGlyLeuTyrAla 1127
3964 CGATACCGCGCGGTTTCCGGGATTCGGCATCGACCGCATCGCGCGC 4013
1128 ThrTyrAla.....AspAspLysSerArgAsnGlyAl 1139
4014 AACGCGCTATTTCGTCAAAAAGCGATACCGCTACGAAAGCGTAAATA 4063
1139 a.....TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsnThrV 1154
4064 TCGCCACCGCGCGCTTCATTCACCGCTACCGC..... 4098
1154 aLysGlyAspAspLeuGlnSerGlnSerTyrLysSerLysGlyPheThr 1170
4099 GCGGCGATTAAAGCAGATTATTCATTCAAACGCGCGCACACATTCGC.. 4146
1171 AlaserLeuGlnAlaGlyTyrLysHisLysLeuAlaGlnPheAsnGlySe 1187
4147 .....ATCAGCGCTTATTTCAGCGCTGCTCT 4171
1187 rGlnGlyThrArgAsnGlnTyrTyrValGlnProGlnAlaGlnValThrT 1204
4172 ATACCGATGCC..... 4182
1204 rPmetGlyValLysAlaAspLysHisArgGlnSerAsnGlyThrLeuVal 1220

```

```

4183 .....GCTTCGGGCAAAAGTCCGACACCGCGTCAATACCGCGTATT 4223
1221 HisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysThrTr 1237
4224 GCGTCAG.....GATTCGCGCAAAACCGC..... 4248
1237 PLeuLysSerHisHisLysMetAspAspGlyLysSerArgGlnPheGlnP 1254
4249 .....AGTCCGAATGGGCGGTAAACGCCGAATCAAAAGGTTTCAG 4290
1254 roPheValGlnValAsnTrpLeuHisAsnSer.....LysAspPheSer 1268
4291 CTGTCCCTCCAGCGCTGCCCGCGCAAAAGCGCGCATGTGACGCGCAACA 4340
1269 ThrSerMetAspGlyValSerVal.....ThrGlnAspGlyAlaArgAs 1283
4341 CACGCGCGCATCAATTTAGGC 4362
1283 nIleAlaGlnIleLysThrGly 1290

```

seq_name: p1r2:E85524

```

seq_documentation_block:
Probable beta-barrel outer membrane protein 20402 [imported] - Escherichia coli (stra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001
C:Accession: E85524
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; DiMantola, E.; Potlamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206531
A:Accession: E85524
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <STOs>
A:Cross-references: GB:AE005174; NID:912513130; PIDN:AA654657.1; GSPDB:GN00145; UMGF:
A:Experimental source: strain O157:H7, substrain EDL933
A:Genetics:
A:Gene: 20402

```

```

alignment_scores:
Quality: 325.00      Length: 1474
Ratio: 0.509
Percent Similarity: 43.284      Percent Identity: 20.692

```

alignment_block:

US-09-303-518D-649 x E85524 ..

Align seg 1/1 to: E85524 from: 1 to: 1349

```

712 TCAGTGGTGGCACAGTCACCTTAGTAGTGAATAAATTAACATAGCCG 761
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
118 SerGlyGlyGlyLeuAspIleThrGlyGlyValThrIleAsnGlyP 134
762 ATATGGTTTTTACCACAGAGGCGTCATT.....GGGACAGTG 802
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134 OleuAsnPheLeuLeuLysGlyThrGlyPheLeuAsnValSerAsnAlaG 151
803 GCTACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAAATTAT 852
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151 LysGlnLeuTyrAlaAspAspLeuTyrGlnSer.....Asn 163
853 GGGTATTGGCAAAAGGCGACCCCTATATAGCAAAAAGCAATGGCTTCCA 902
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 SerGlyMetArgHisAspArgGlyTyrPheAsnValSerAsnGlyLys 180
903 GCTG..... 906
180 sIleHisValLysGlyThrSerArgLeuThrTyrLeuGlnGlyAsnValS 197
907 .....GTTCTAAAGATTGGTTCTATGATGAATAATCTTT 939

```

```

197  erglgluglyserglnvalasnerglthrphemetglvaltyr. 213
198  .....
199  940  GGTGAGATACCATTCAGTATTCTAGAACACGTCMAATGGGAATA 989
200  .....
201  214  .....glyserly 216
202  .....
203  990  CTCTTTTAC.....GACGATAATATGCGACAGAAAAATCAATG 1030
204  .....
205  216  rglglglnsnglnltyrleuservalasnsngly...Glyglvalasna 232
206  .....
207  1031  CCAACATGTGACACAAAT..... 1047
208  .....
209  232  laarglsglnlleserleuglytyrtyraspglnvalserasplthrthr 248
210  .....
211  1048  .....TCTGCTTAATAGATTAAACACAGAACGTTCAATG.. 1086
212  .....
213  249  leuvalasergluglylysileseralaprothrleuserleuse 265
214  .....
215  1087  .TTTAATGTTTCTTATCCGAGACAGACAACTGTTATCATGCTG 1135
216  .....
217  265  rthraserleuvalaleuglyalaglnglyseralalaysa 282
218  .....
219  1136  CAGGTGTGTACACAGTTATCGACCACTGAATAATGAGAAATATT 1185
220  .....
221  282  laalagllyle.....leasplagllyle 291
222  .....
223  1186  TCCTTTATGACGAGAAAGCGAATTGACTTACGACCAAC..... 1230
224  .....
225  292  gluphevaltrpalalysthrserglulyslthrleuasnsth 308
226  .....
227  1230  ..... 1230
228  .....
229  308  rasplyasplathrilleseralaspllevalserglyserylulyl 325
230  .....
231  1231  .....ATCAATCAAGGTCTGAGGATTATTTCCAGAGGATT 1272
232  .....
233  325  euglytyrilleasnhalaleuasnlythrthrlyleuthrlyasp... 340
234  .....
235  1273  ACGGTCTGCTGAAATACGAAACTGGCAAGCGCGGCTTCATAT 1322
236  .....
237  341  .....Asnseralapheserlylys...Vallysl 350
238  .....
239  1323  CAGTGAAGACATACCGTT.....ACTTGAAG 1351
240  .....
241  350  eglglnsnnglyalaleuglylthrghlnsnllegllythralaglthr 367
242  .....
243  1352  TAAACGGCGGTGCAACGACCGCTGTCCAAATCGCGCAAGGACGCTG 1401
244  .....
245  367  leasn.....Asnargglylysleu 373
246  .....
247  1402  CACGTTCAAGC.....AAGGGGAAACCA 1427
248  .....
249  374  histleuylsalaspspsermethrphelalasnlylleserglyas 390
250  .....
251  1428  AGGCTCGATCAGCGTGGCGAGCGGTACAGTCATTTGGATCAGACGCGAG 1477
252  .....
253  390  ngllythrilleserlleaspserylthrvalgluleu..... 402
254  .....
255  1478  ACGATAAAGGCAAAAACAGCTTACT...GAATCGGCTGGTGCAC 1524
256  .....
257  403  .....Thrglyasnasnlyralapheserlytyrilleasplvalaser 417
258  .....
259  1525  GGCAGG.....GCTAGCGTCAACT 1544
260  .....
261  418  glvalalavalalalalleserlunasplysasnllegllyarghalaglle 434
262  .....
263  1545  GAATGCCGATTAATCAGTTC.....AACCCGACAAACTCTATT 1582
264  .....
265  434  uaspyalaspglylysleuglnlleasnhalasnlysasptrvalphea 451
266  .....
267  1583  TCGGCTTTCGCGCGGACGT.....TTGGATTAAAGCGGCAAT 1620
268  .....

```

```

451  spasnaspheuglulgllyargglyllevalgluleasnmethglyasnhs 467
452  .....
453  1621  TCGCTTGTGTTCCACCGTATTCMAAATTCGAT.....GAAGGGCGCAT 1664
454  .....
455  468  glupheserlphespluphealatythrthrasptrphenglnglyserle 484
456  .....
457  1665  GATTGTCAACCAATCAATCAAGACMAAAGATCCACCGTTACATTACAGGCA 1714
458  .....
459  484  ualapheglinsn..... 488
460  .....
461  1715  ATAAAGATATTGCTTCAACCGCGCATATACACAGCTTGATAGCAAAAA 1764
462  .....
463  489  .....Thrrhrpheasn.....Leugllysasna 497
464  .....
465  1765  GAATTTGCTTACACAGTGTGTTTGGCGAGAAAGATACGACCAAAACGAA 1814
466  .....
467  498  glupheleuglnlysgly.....glylthrhalagllyl 509
468  .....
469  1815  CGGGCGGCTCAACCTGTTTACACGCGCGCGACAGACCGCACCTTC 1864
470  .....
471  509  ngllyserleuvalthrvalgllysglyalanhlsertleerthrleug 526
472  .....
473  1865  TCGTTTCGCGGCGGACAAATTTAAACGCAACATCACG..... 1902
474  .....
475  526  lypheaserlygllythrvalasplhegllyalaleuthrhalagllyaln 542
476  .....
477  1903  CAACAAACGCGCAACCTGTTTTCAGCGGACAGACACGACGCGCCTA 1952
478  .....
479  543  methrghlgllythrvalasnvalser..... 551
480  .....
481  1953  CAATCATTTAAACGACATTGTGCGCAAAAGAGGCAATTCCTCGCGGG 2002
482  .....
483  552  .....Lysthrleuaspheuarlylg 559
484  .....
485  2003  AA.....ATCGTGTGACAAACGACTGATC..... 2028
486  .....
487  559  luglyvalilleglnvalseraspservalalargvalserarg 575
488  .....
489  2029  .....AACCGCAC 2036
490  .....
491  576  asplleaspservalaleuserleuthrghlvalasplaspglyasnserthr 592
492  .....
493  2037  ATTATAA.....CGGAAA 2050
494  .....
495  592  rtlelyslleuvalasplaglnclyalagluvalleuglyasplagllya 609
496  .....
497  2051  ACTTCCAAAT.....AAGGCGGACGCGGTGTTCCGCAATGTT 2094
498  .....
499  609  snleuglnleuglnasplysasnglylnlleuserserseralagln 625
500  .....
501  2095  GCCAAAGTCAAGCGCATTTGGCATTTGACCAATCACGCCCAAGCGTTT 2144
502  .....
503  626  argasplleglin.....glnasnlyglinlyslalalava 637
504  .....
505  2145  TGGTGTGCGCCGACATCAACGACCAAAATCTGTACAGCTGGAGTGA 2194
506  .....
507  637  lgllythrtyraspyrtyrargleu...Thrserylvalasnaasnspglyl 653
508  .....
509  2195  CG.....GCTTGCACAAATGT...GTGCAAAAACCAATTAC 2229
510  .....
511  653  eutyrllegllytyrlyleuthrghlnleuaspheuhlsatthrpsrser 669
512  .....
513  2230  GACGATTAAGTGAATGCTTCATTTGACTAAGACCGAC..... 2265
514  .....
515  670  asplaleuvalaleuserserasnlylylserlgluasnalalalasple 686
516  .....
517  2266  .....ATCAGCGCAATGTCATCTTGC..... 2289
518  .....
519  686  useralalysllethrlyserglyaspleualapheserleuglnysg 703
520  .....
521  2290  .....GATCAGCTCATTTAAATTCACA 2313
522  .....
523  703  lyglnthrvalserleuserasnlysaspsasnsplythrghlvalthr 719
524  .....

```

```
2314 .....GGGCTTGCACACTCAAC.....GGCAA 2336
|||
720 AspleuargSerGlyThrLeuLeuLeuAsnAspAsnValLeuGlyAs 736
2337 T.....CTTAGTCGAATGCGC 2353
|||
736 nThrHisGluLeuArgLeuAlaIaGluThrGluLeuAspMetAsnGly 753
2354 ATACAGCTTATACAGTCAGCCACACGCCCAACCAAGCGCAACCTTAGC 2403
|||||
753 IAspGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeuLeuSer 769
2404 CTCGTGGCAATGCCCAAGCAACATTTATATCAAGCCACATTAAGCGCAA 2453
|||
770 LeuAsnGlySerLeuThrValThrAsnGlyGlyThrSerThrGlySe 786
2454 CACATCGCGCTTCGGGCAATGCTTCATTTAATCTAAGCGCACCGCGTAC 2503
|||
786 rLeuThrGlySerGlyLeuLeuAsn.....IleG 796
2504 AAAACGGCAGCTGTGACGCTTCCGGCAACGCTAAGCAACGTAAAGCAT 2553
|||
796 lngGlyThrLeuAspIleAlaGlyAsp.....Asn 806
2554 TCCGCACTCAACGTAATGTCTCCCTAAGCGGATAGGCA.....GT 2594
|||
807 SerAsnLeuThrAlaAsnValAlaIaAsnSerAlaAsnValLeuVa 823
2595 ATTCCATTTTGAAGACAGCGCGCTTACCGGACAAATCAGCGGCGGCAAG 2644
|||
823 IserHisAlaGlnGly..... 828
2645 ATACGGCATTTACCTTAAAGACAGCAATGAGACGCTGCCGTACAGCAG 2694
828 ..... 828
2695 GAATTAGGCAATTTAAACCTTGAC...AACGCCACATTACACTAATTC 2741
|||||
829 ...LeuGlySerAlaAsnValGluAsnAsnGlyThrLeuAlaLeuAsnAs 844
2742 CGCCTATGCCACGATGCGGAGG.....GGGCAACCGGCA 2779
|||
844 nSerAlaGluIuysArgAlaAlaIaSerValAsnGlyAlaLeuGlyAla 861
2780 GTCCGACAGATGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTA 2829
|||
861 snLeuThrAsnAsnGly..... 866
2830 TTATCGGTTACACCGCCACTTCGGTGAATCCGTTTCAACAGCTGAC 2879
|||
867 ...ThrLeuMetThrGlyMetSerGlyGlnAlaGlyAsnValLeuVa 882
2880 GGTAAAGCGCAATTGACCGT...CAGGGAATTCCTCGTTATGTCGG 2926
|||
882 IValIysGlyAsnGlyHisGlyAsnAsnGlyGlnLeuValMetAsnThr 899
2927 AACTCTGGGTCACCGCAGC.....GACAAATTGACGTGGCGGAAAGT 2970
|||
899 alLeuAsnGlyAspAspSerValThrAspIysLeuValValGluGlyAs 915
2971 TCCGAAGGCACTTACACTTGGCGGTCAACATACCGGCAAGCAACCTGAC 3020
|||
916 ThrSerGlyThrThrAlaValThrValAsnAsnAlaGlyGlyThrGlyAl 932
3021 AAGCCTC.....GAACAATTGACGTAAGTGAAGAAAGCAACA 3061
|||
932 aIysThrLeuAsnGlyIleGluLeuIleHisValAspIysSerGluG 949
3062 AACCGCTGTCCGAACACCTTAATTCACCCGTCCAAAGCAACAGCTGCAT 3111
|||
949 IyGluIphe.....ValGlnAlaGlyArgTrileVal 958
3112 GCCGGCGCGTGGCTTACCACTCATCCGCAAAAGACGGCGAGTCCGCT 3161
|||||
959 AlaGlyAlaIaIysAspIyThrLeuAlaIaArgGlyGlnGly..... 971
3162 GCATATTCGGGTCAAGAAACAAGAGCTTCCGACAAACTCGGCAAGGCAG 3211
971 ..... 971
3212 AAGCCAAAAAAGAGCGGCAAAAAAGACACGCGCAAGCCTTGACGGCTG 3261
972 .....AlaAsnSerGlyAsnTrpIy 978
3262 ATTGCGCGCGCGCGCGATCCGTCGAAAGACAGAAAGCGTTGCCAACC 3311
|||
979 LeuThrSerGlySerAspSerProGlu..... 987
3312 GCCCGCGCAGGACAGCGGGAATGTGGCATTTATGACAGCGGAGAGAG 3361
987 ..... 987
3362 AAAAAAAGCGGTGACAGCGGATAAAGACACCGCCTTGCGAAACAGCGC 3411
988 .....LeuGlnPro 990
3412 GAAGGGAAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGC 3461
|||
991 GluProAspPrometProAsnProGluProAsnProAsn..... 1003
3462 CCGCGGGGATTTCGCGCACTGACACCCCAACCGGACCGGACCGGACG 3511
|||
1004 .....ProGluProAsnProAsnProIleProIleProGly 1016
3512 GCGACCTG.....ATCAGCGCTTATGCCAATAGCGGTTG 3546
|||
1016 roAspLeuAsnValAspAsnAspLeuArgProGluAlaGlySerTrile 1032
3547 AGTGAATTTTCGCGCAGCTCAACGAGCTTTCGCG.....GTAACAGA 3590
|||||
1033 AlaAsnLeuAlaIaIaIaIa...AsnThrMetThrThrArgLeuHisGly 1048
3591 CGAATTA.....GACCGGCTATTTGCCGAAAGACCGC 3622
|||
1048 uArgLeuGlyAsnThrTrpIyThrAspMetValThrGlyLeuGlnIysG 1065
3623 GCAACGCGGCTTGGACAGCGGCAATCCGGGACACCAACACTACCTTCG 3672
|||
1065 InThrThrMetTrpMetArg.....HisGluGlyGlyHisAsnIysTrp 1079
3673 CAAGATTTCCGCGCTACCGCCAAACCAACGACCTGCGCAATCGGTAT 3722
|||
1080 ArgAsp..... 1081
3723 GCAGAAAAAAGCTCGGACGCGCGCGCTGCGCATCTGTTTCGCAACAC 3772
|||
1082 .....GlySerGlyGlnLeu.....LysT 1088
3773 GGACGAAACACCTTCAGACGCGCATTCGCAACCTGGGACGCGCTTGGC 3822
|||
1088 hrGlnSerAsnArgIyValLeuGlnLeuGlyGlyAspVal..... 1101
3823 CACGCGCGCTTTCGGGCAATACGCGCATGACAGGTTCTACATCGGCAT 3872
|||
1102 .....AlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValGlyVa 1116
3873 CAGCGCGGCGCGGCTTTAGC...ACCGGACCGCTTTCAGAGCGCATCG 3919
|||
1116 lMetAlaGlyTrpIyAsnSerAspSerIyThrIleSerSerArgTrhg 1133
3920 GAGCAAAATCCGCGCGCGCGCTGCTGCATAC.....GGCATTCAGGCA 3963
|||
1133 IyTrArgAlaIyAsAlaSerValAsnGlyTrpSerThrGlyLeuIyAla 1149
3964 CGATACGCGCGGCTTTCGGCGGATTCGGCATCGAAACCGCATCGCGCGC 4013
```



```

3834 TTTC.....GGGCAATACGATCGACGATTC.....TACA 3865
      |||||.....:|||||
768 yphetylLeuAspGlyValValLysLeuAsnArgPheLysSerAsnVala 785
      |||.....:|||||
3866 TCGGCATACAGCGCGCGCGGTTTTCAGCAGCGACCTT..TCAGAC 3912
      |||.....:|||||
785 IagLysMetSerSerGlyValAlaAlaAsnGlySerTyHisSerAsn 801
      |||.....:|||||
3913 GGCATCGAGAGCAAAATCCGCCCGCGCTGCTCATTCAGCGCATTCAGG 3962
      |||||.....:|||||
802 GlyLeuGlyGlyHis.....:|||||
3963 ACGATTACCGCGCGGTTTCGCGGATTCGCGATCGACAAACGACATCGCG 4012
      |||.....:|||||
809 rGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerI 826
      |||.....:|||||
4013 CAACGCGCTATTCGTCCAAAAAGCGGATTACCGCTACGAAAC..... 4056
      |||.....:|||||
826 eutHrGlyPheThrAlaAspAsnProGlyTyHisLeuSerAsnGlyMet 842
      |||.....:|||||
4057 .....GTCAATATCGCCACCGCGCGCTTCGATTCAAACGCTACCGC.. 4098
      |||||.....:|||||
843 LysSerLysSerValAspThrArgSerIle.....TyrArgI 855
      |||||.....:|||||
4099 ..GCGGCGATTAAAGCAGATTATTCATTCAAAACGCGCGCAACACATTCCA 4147
      |||.....:|||||
855 uLeuGlyAlaThrLeuSerTyHisMetArgLeuGlyAsnGlyMetIuV 872
      |||.....:|||||
4148 TCAGCGCTATTTCAGCGCTGCTGCTTACCGATGCGCGCTTCGGGCAAGTC 4197
      |||.....:|||||
872 aLgIuProTyrLeuLysAlaAlaValArgLysGluPheValAspAsn 888
      |||.....:|||||
4198 CGAACACGCGCTCAATACCGCGCTGCTGCTGCGATTC.....GSCAA 4241
      |||||.....:|||||
889 ArgValLysValAsnSerAspGlyAsnPheValAsnTyLeuSerGlyAr 905
      |||||.....:|||||
4242 AACCGCGATCGCGATGCGCGGCTAAACGCGCAAAATCAAGCTTCACGC 4291
      |||.....:|||||
905 gArgGlyTyLeuTyGlnAlaGlyLeuLysAlaSerPheSerSer...ThrL 921
      |||||.....:|||||
4292 TGTCGCTCCAGCGCTGCGCGCGCAAGCGCGCAACTGGAAGCGCAACAC 4341
      |||||.....:|||||
921 euSerGlyHisLeuGlyValGly.....TySerHis 931
      |||||.....:|||||
4342 AGCGCGGCGCATCAAA.....TTAGGCTACCGCTG 4371
      |||||.....:|||||
932 SerAlaGlyValGluSerProTyrAsnAlaValAlaGlyValAsnTyr 947

```

seq_name: p1r2:E95206

seq_documentation_block:
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain T1

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: E95206

R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Omayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,

son, T.; Hickey, E.R.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: E95206

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4776 <RUR>

A:Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:914973269; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

alignment_scores:
Quality: 311.00 Length: 1067
Ratio: 0.519 Gaps: 31
Percent similarity: 56.139 Percent identity: 20.806

alignment_block:
US-09-303-518D-649 x E95206 ..

Align seg 1/1 to: E95206 from: 1 to: 4776

```

1273 ACGGCTGCTCGCTGAATAATACGAACTTGGACGCGCGCGCTTCATAT 1322
      |||.....:|||||
103 ThrSerAlaSerGluSerAlaSerThr.....SerAlaSerThrSerI 117
      |||.....:|||||
1323 CAGTGAAGACAGTACCGCTTACTTGAAGTAAGTAAAGCGCGTGAACGAC 1372
      |||||.....:|||||
117 eSerAlaSerSerThrValValGlySerGlnThrAlaAlaIaThrGluA 134
      |||||.....:|||||
1373 GCGTGTCCAAATTCGCAAGACGCGTGCACGTTCAACGCAAGGGGA 1422
      |||||.....:|||||
134 IaThrAlaLys.....LysValGluGluAspArgLys 144
      |||||.....:|||||
1423 AACCAAGGCTCGATCAGCGTGGCGACGCTACATGCTATTCAGTACGAC 1472
      |||||.....:|||||
145 LysProAlaSerAspTyrValAlaAlaSerValThrAsnVal...AsnLeuG 160
      |||||.....:|||||
1473 GCGACAGCATTAAGCAAAACAAACAGCC..... 1500
      |||||.....:|||||
160 nSerTyAlaLysArgArgLysArgSerValAspSerIleGluGlnLeuL 177
      |||||.....:|||||
1501 .....TTTAGTAAATCGGCTTGCTGACG 1524
      |||||.....:|||||
177 euAlaSerIleLysAsnAlaAlaValAlaPheSerGlyAsnThrIleValAsn 193
      |||||.....:|||||
1525 GCGAGGCGTACGCTGCACTGATGATCCGCTATTCAGTCAACCGGACAA 1574
      |||||.....:|||||
194 GlyAlaProAlaIleAsnAlaSerLeuAsnIleAlaLysSerGlnThrLy 210
      |||||.....:|||||
1575 ACTCTATTCGCGCTTCGCGCGGACGTTTGATTTAAACGCGCATTCGC 1624
      |||||.....:|||||
210 sValTyL.....ThrGlyGluGlyValAspSerValTyArgValP 224
      |||||.....:|||||
1625 TTTGCTTCACCGTATTCAAATACCGATGAAGGGCGATGTTTGCAAC 1674
      |||||.....:|||||
224 rolleTyTyTyLysLeuLysValThrAsnAspLysSerLysLeuThr... 239
      |||||.....:|||||
1675 CACAATCAAGCAAGAATCCACCGCTTACCATTAACA...GGCAATTAAGA 1721
      |||||.....:|||||
240 .....PheThrTyThrValThrTyValAsnProLy 250
      |||||.....:|||||
1722 TATTGCTACACCGCATATACACAGCTTGATAGCAAAAAGAAATTG 1771
      |||||.....:|||||
250 sThrAsnAspLeuGlyAsnIleSerSerMetArgProGlyTySerIleT 267
      |||||.....:|||||
1772 CCTRACAAAGGTGG..... 1785
      |||||.....:|||||
267 yTrAsnSerGlyThrSerThrGlnThrMetLeuThrLeuGlySerAspLeu 283
      |||||.....:|||||
1786 .....TTTGGCGAAGAAAT...ACGACCAAAAAGACGCGGCGCT 1823
      |||||.....:|||||
284 GlyLysProSerGlyValLysAsnTyIleThrAspLysAsnGlyArgI 300
      |||||.....:|||||
1824 CAACCTTGTTTACAGCGCGCGCGCAAGACCGCACCTGCTGCTTCGCG 1873
      |||||.....:|||||
300 nValLeuSerTyAsnThrSer.....ThrMetThrThrGlnG 313
      |||||.....:|||||
1874 GCGGAACAAATTTAAACGCAATCAGCAACAAACGCAAACTGTTT 1923
      |||||.....:|||||
313 LysSerGlyTyThrThrPbLysAsnGlyAlaGlnMetAsnGly.....Phe 327
      |||||.....:|||||
1924 TTCAGCGGACAGCAACACCGCGCTTACATATTTAAAGACATTTG 1973
      |||||.....:|||||
328 PheAlaLysLys.....GlyTyArgLy...LeuThrSerSerTr 339

```



```
1974 GTCGCAAAAAGAGGGCATTCCT...CGGGGGAATGCTGTGCACACG 2020
      |||
      |||
339 pThr.....ValProIleThrGlyThr.....AspIris 349
2021 ACTGATACACCGCAGCATTTAAAGCGGA.....AAC 2052
      |||
      |||
349 erpThrPheThrProTyrAlaIleThrAspArgIleGlyIleAsn 365
2053 TTCGCAATTAAGCGGACAGCGGTGTTCCCGCAATGTTGGCAAGT 2102
      |||
      |||
366 TyrPheAsnGlyGlyValValGluSerThrThrSerGln.. 381
2103 GAAAGCGGATTTGCATTGAGCAATCACGCCCAAGCAGTTTGGTCTG 2152
      |||
      |||
382 .....SerLeuSerGlnSerLysSerLeuSerValSerIAs 394
2153 CACGCGATCAAGCCACACATCTGTACAGCTTGCAGCTGGACGGCTCTG 2202
      |||
      |||
394 erGlnSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer 410
2203 ACAATTTGTGCAAAAACCATTTACCGCAGATTAAGTGAATGCTTCAT 2252
      |||
      |||
411 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer.. 426
2253 GACTAAGACCGACATCAGCGGCAATGTCATTTGCCGATCAGCTCAT 2302
      |||
      |||
427 .ThrSerAlaSerValSerAlaSerThrSerAlaSerAlaSerIAs 443
2303 TAAATCTCACAGGGCTTGCACACTGCAGCAATCTTACTGCAATGCG 2352
      |||
      |||
443 hrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerIAsThr 459
2353 GATACAGCTTATACAGTCACGCCAACGCCCAAAACGCAACCTTAG 2402
      |||
      |||
460 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 476
2403 CTCGCGGCAATGCCCAAGCAACATTTAATCAGCCACATTAACGCGA 2452
      |||
      |||
476 rAlaSerAlaSerAlaSerThrSer.....AlaSerGluSerIAs 490
2453 ACACATCGGCTTGGGCAATGCTTCATTTAATCTAAGCGACAGCGCGTA 2502
      |||
      |||
490 erThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerIAs 506
2503 CAAAAGCGCAGTGTACGCTTCCGCAACGCTAAGCAACGTAAGCCA 2552
      |||
      |||
507 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAs 523
2553 TTCGCGATCAAGCGTAAATGTCCTTACCGCATAGCAGATTCATT 2602
      |||
      |||
523 rSerAla...SerGlySerAlaSerThrSerThrSerAlaSerIAs 539
2603 TTGAAAGCAGCGGCTTTACCGGACAATAGCGGCGCAAGGATACGGA 2652
      |||
      |||
539 hrSerAlaSerAla...SerAlaSerThrSerAlaSerAlaSerIAs 554
2653 TTAACTTAAGACAGCAATGACGCTGCCCTAGCAGCAGGAATTAGG 2702
      |||
      |||
555 lIeSerAlaSerGluSerAlaSerThrSerAlaSerGluSerIAs 571
2703 CAATTTAACTTGTGACACGACCATTAACCTAATCCGCTATTCGCC 2752
      |||
      |||
571 rSerThrSerAlaSerAlaSerThrSerAlaSerGluSerIAsThr 588
2753 ACAGTGGCGGAGGGCGCAACGCGCAGTGCAGAGATGCGCGCGCGCG 2802
      |||
      |||
588 erAlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 604
2803 CGTTGCGCGCGCTTCCGCTTATTCCTTATTCCTTACCGCCCACTTC 2852
      |||
      |||
605 AlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 621

2853 GTTAGAATCCCTTTTAAACGCTGACGGTAAACGCAATTTGAACGCTC 2902
      |||
      |||
621 uSerAlaSer..ThrSerAla.....SerAlaSerAlaSerThr 633
2903 AGGGAACATTCGCTTTATGTGCGAACCTCTGCGCTACCGCAGCGCAAA 2952
      |||
      |||
634 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 650
2953 TTGAGCTGGCGGAAAGTTCCGAGGCACTTACCTTGCGGTCAACA 3002
      |||
      |||
650 rAlaSerAlaSerAlaSerThrSerAla.....SerValS 662
3003 TACCGCAACGAAACCTGCAAGCTTGCAGCAATTTGACGATGTGAGGA 3052
      |||
      |||
662 erAlaSerThrSerAlaSerAlaSerAlaSer.....Thr 673
3053 AACACACAAACCGCTGTCCGAAACCTTAATTCCTGCAAAACGAA 3102
      |||
      |||
674 SerAlaSerAlaSerAlaSerThrSerAlaSerGluSerIAsThr 690
3103 CACGTGATCGCGCGCGCGCTTACCACTATCCGCAAGAGCGGA 3152
      |||
      |||
690 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 707
3153 GTTCCGCTTCATTAATCCGTCAAAGAACAGCTTTCCGCAACTCG 3202
      |||
      |||
707 lSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThr 723
3203 GCAAGGCAAGACCAAAACAGCGGCAAAAGACAGCGCAACCTT 3252
      |||
      |||
724 Ser..AlaSerAlaSerThrSerAlaSerGluSerIAsThr 739
3253 GACGGCTGATTTGCGCGCGCGCGATCCGTCGAAAGACAGAAAGCGT 3302
      |||
      |||
740 SerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 756
3303 TGCCGAACCGCGCGCGGAGCGAGCGGGAATGTCGGCATTTGCGAG 3352
      |||
      |||
756 rAlaSerAlaSerThrSerAlaSerAla..... 765
3353 CGGAGGAAGAGAAAAACGGGTGACGCGGATTAAGACCGCGCTTGGCG 3402
      |||
      |||
766 .....SerAlaSerThrSerAlaSerAlaSerIAsThr 778
3403 AAACGCGCGAAGCGGAAACCGCGCGCTACACCGCTTCCCGCGCGC 3452
      |||
      |||
779 AlaSerGluSerAlaSerThrSerAlaSerAlaSerIAsThr 795
3453 CCGCGCGCGCGCGGATTTGCGCAACTGCAACCCCAACCGCAGCGCC 3502
      |||
      |||
795 lAsSerAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 811
3503 AACCGCAGCGGACCTGATCAGCCGTTATGCAATAGCGTTGAGTAA 3552
      |||
      |||
812 SerGluSerAlaSerThrSerAlaSerAlaSerIAsThr 828
3553 TTTCCGCGACGCTCAACAGCGCTTTCGCGCTACAGACGAATTAAGCC 3602
      |||
      |||
828 rAlaSerAlaSerThrSerAlaSerAlaSerIAsThr 845
3603 CGTATTTGCGGAGACCGCGCAACGCGGTTTGACAAAGCGGATCCGG 3652
      |||
      |||
845 lAsSerAlaSerThrSerAlaSerAlaSerIAsThr 861
3653 ACACCAAAACACTACCGTTCCGCAAGATTTCCGCGCTTACCGCAACAA 3702
      |||
      |||
862 SerAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerIAs 878
3703 GACCTGCGCAAAATCGGTATGCAAAAAACCTGCGGACGCGCGCTCG 3752
      |||
      |||
878 rAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerIAs 895
3753 CATCTGTTTTCGACACCGGACCGAAACACCTTTCGACGAGCGGATCG 3802
```

```

|||||  |||:|||||  ::  |||:|||||  |||:|||||
1895  laser.....ThSerAlaserAlaserThSerAlaser 908
3803  GCAACGCGGACGCGCTTGGCCAGCGCGCTTTTGGGCAATACGCGCATC 3852
|||:|||||  |||  |||:|||||  |||  |||:|||||  |||
909  AlaserAlaserThSerAlaserAlaserThSerAlaser 925
3853  GACAGGTTTACATCGGCATACGCGCGCGGCTTTAGCAGCGGACG 3902
|||:|||||  |||:|||||  |||  |||:|||||  |||
925  aSerAlaserThSerAlaserAlaserThSerAlaser 942
3903  CCTTCAGCGCGCATCGGCAAAATCGCGCGCTTGCATGATACG 3952
|||:|||||  |||:|||||  |||:|||||  |||:|||||  |||
942  erThSerAlaserThSerAlaserThSerAlaser 958
3953  GCATTGACGACGATACCGCGCGCTTGGCGGATTCGGCATGACG 4002
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
959  ThSerAlaserAlaserThSerAlaserAlaser 975
4003  CACATCGCGCGACGCGCTATTGCTCCMAAACGGATTAACGCTACGA 4052
|||:|||||  |||  |||:|||||  |||  |||:|||||  |||
975  rSerAlaserThSerAlaserThSerAlaserAlaser 992
4053  AAACGTCATATCGCACCCCGCGCTTGCATTCACCGCTACCGCGCG 4102
|||:|||||  |||  |||  |||  |||  |||  |||  |||  |||
992  erAlaserAlaserThSerAlaserThSerAlaserThSer 1008
4103  GCATTAGCGCAGATTATTCATTCAACCGCGCGCACATTCATTCACG 4152
|||  |||:|||||  |||  |||  |||:|||||  |||
1009  Ala..SerAlaserThSerAlaserThSerAlaser 1024
4153  CCTTATTGAGCGCTGTCTATACGATGCGCTTGGGCAAGTCCGAC 4202
|||  |||  |||  |||  |||  |||  |||  |||  |||
1025  AlaserGlaserAlaserThSerAlaserAlaserThSer 1041
4203  ACGCGTCATATACCGCGCTATTGCTCAGAGATTGGCAAAACCGCGACTG 4252
|||:|||||  |||:|||||  |||:|||||  |||:|||||  |||
1041  aSerAlaserAlaserThSerAlaserAlaserThSer 1058
4253  CGGATGGGCGCTAAGCGCGCAATCAAGGTTTCAGCGTCCCTCCAC 4302
|||:|||||  |||  |||  |||  |||  |||  |||  |||
1058  erGlaserAlaserThSerAlaserAlaserThSer 1074
4303  GCTGCCGCGCGCAAGCGCGCACTGGAAGCGCAACACAGCGCG 4347
|||:|||||  |||  |||  |||  |||  |||  |||  |||
1075  GlaserAlaserThSerAlaserAlaserThSer 1089

seq_name: pir2:A47675

seq_documentation_block:
68k outer membrane protein P.68 pertactin - Bordetella bronchiseptica
C:Species: Bordetella bronchiseptica
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A47675
R:Li, J.; Fairweather, N.F.; Novotny, P.; Dougan, G.; Charles, I.G.
J. Gen. Microbiol. 138, 1697-1705, 1992
A:Title: Cloning, nucleotide sequence and heterologous expression of the protective oute
A:Reference number: A47675; MUID:92407514
A:Contents: CN7531
A:Accession: A47675
A>Status: Preliminary
A:Molecule type: nucleic acid
A:Residues: 1-911 <L11>
A:Cross-references: GB:X54815; GB:S46416; NID:g39396; PID:CAA38584.1; PID:g39397
A>Note: sequence extracted from NCBI backbone (NCBIN:113318, NCBI:P.113319)

alignment_scores:
Quality: 295.50 Length: 1079
Ratio: 0.612 Gaps: 50
Percent Similarity: 44.764 Percent Identity: 21.965

alignment_block:

```

```

US-09-303-518D-649 x A47675 ..
Align seg 1/1 to: A47675 from: 1 to: 911
1417 GGGGAAACCAAGCTGCATACGCTGGCGACGCTACAGTCATTTTGA 1466
|||:|||||  |||:|||||  |||  |||:|||||  |||
45  GlyIuArGlnHisIle..... 53
1467 TCAGAGGCGACGATTAAGCGCAAAACGCTTAGTGAATCGGCT 1516
|||:|||||  |||  |||  |||  |||  |||  |||  |||
54  .LysGlnSerAspGlyAlaGlyAlaIArgThrAlaThrGlyThrIle 70
1517 TGGTCAGCGCGAGGGTACGCTGCACTGAATCCGATATCATCTAC 1566
|||:|||||  |||  |||  |||  |||  |||  |||  |||
70  yValSerIArgIuAlaGlnGlyValIleuLeuGluAsn..... 83
1567 CCCGCAAACTATTTGCGCTTGGCGGCGAGCTTGATTTAAAGG 1616
|||:|||||  |||  |||:|||||  |||  |||:|||||  |||
84  ..ProAlaIaGluLeuArgPheGlnAsnGlySerValThrSerSer 99
1617 GCATTGCTTGTCTCCACCGCTATTCAAAATACGATGAAGGCGATGA 1666
|||  |||  |||  |||  |||  |||  |||  |||  |||
99  Y.....GlnLeuPheAspGlyVal.... 106
1667 TTGTACACCAATCAAGCAAAAGATTCACGCTTACCTTACAGCAAT 1716
|||:|||||  |||  |||:|||||  |||  |||:|||||  |||
107  .....ArgArgPheLeuGlyThrValThrValIuLysAlaGly 118
1717 AAA.....GATATGCTACACCGCAATACACAGCTTGATAG 1757
|||  |||  |||  |||  |||  |||  |||  |||
119  LysLeuValAlaAspHisAlaThrIleuAlaAsn..... 129
1758 CAAAAAAGAAATTCCTACACGCTTGTGGCGGAAAGATACAGCA 1807
|||:|||||  |||  |||  |||  |||  |||  |||  |||
130  .....ValSerAspThrArg 135
1808 AAACGACGCGGCGCTACACCTTGT.....TACGACCGCGCG 1845
|||:|||||  |||  |||  |||  |||  |||  |||  |||
135  spAspAspGlyIleAlaLeuTyValAlaGlyIuGlnAlaGlnAlaSer 151
1846 GCAGAGACCGCGACCGCTGTCTTCCGGCGGAACAATTTA.....AA 1889
|||  |||  |||  |||  |||  |||  |||  |||
152  IleAlaAspSerThrIleuGlnGlyAlaGlyValArgValGluArgI 168
1890 CGGCACATATCGCAAAACAGCGCAACGCTTTTCAGCGGC..... 1932
|||:|||||  |||  |||  |||  |||  |||  |||  |||
168  yAlaAsnValThrValGlnArgSerThrIleValAspGlyIleuHisI 185
1933 .....AGACCAACACCGCGCTACATCATTTAAACGACCAT 1971
|||:|||||  |||  |||  |||  |||  |||  |||  |||
185  IeGlyThrIleuGlnProLeu..... 191
1972 TGGTCGCAAAAGAGGCAATCTCTCGCGGGAATCTGTGGGACACGA 2021
|||  |||  |||  |||  |||  |||  |||  |||
192  .....GlnProGluAspLeuProProSerArgValValLeuGlyAsp 206
2022 CTGATATCAAC.....CGCACAT 2038
|||:|||||  |||  |||  |||  |||  |||  |||  |||
206  rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValSerValP 223
2039 TTAAAGCGGAACCTTCCAAATTAAGCGGACAGCGGTGTTCCCGC 2088
|||  |||:|||||  |||  |||:|||||  |||  |||:|||||  |||
223  heGlyAlaAsnGluLeuThrValAspGlyHisIleThrGlyIArg 239
2089 AATGTTGCCAAGTGAAGGAGATTGGCATTTGACAAATACCGGCCAAG 2138
|||:|||||  |||  |||  |||  |||  |||  |||  |||
240  AlaAlaGly.....ValAlaAl 245
2139 AGTTTGTGCTGCACCGCATCAAGCACACATCTGTACAGCTTCG 2188
|||:|||||  |||  |||:|||||  |||  |||:|||||  |||
245  aMetAspGlyAlaIleValHisIleuGlnArgAlaThrIleArgArgGly 262
2189 AC..... 2190

```

```

262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyPhe 278
2191 .....TGGACGGCTCTCAACAATTGTCTCGAAAAAC 2222
279 GlyProLeuLeuAspGlyTyrGlyVal..... 288
2223 CATTACGCGAATAAAGTATGCTTCATTGACTAAGACCGACATCAGC. 2271
289 .....AspValSerA 292
2272 ..GGCAATGTGATCTTGGCATCAGCTCATTTAAATCTCACAGGCTT 2319
292 spSerThrValAspLeuAlaGln.....SerIleValGluAla 304
2320 GCCCACTCAAGCAATCTTAGTCAAAATGGGATACAGCTATATCAGT 2369
305 ProGlnLeuGlyAlaIleAlaIleAlaGlyAlaGlyAlaValThrVal 321
2370 CAGCCACAAAC.....GCCACCCAAAACGGCAACTTAGCTCTGGGCA 2413
321 IserGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyG 338
2414 ATGCCCAAGACACATTTAAT.....CAAGCCACATTA 2445
338 IyGlyAlaIleArgArgPheProProAlaSerProLeuSerIleThrLeu 354
2446 AAGCGCAACATCGGCTTCGGGCATGCT...TCATTTAATCTAAGCA 2492
355 GlnAlaGlyAlaIleAlaGlnGlyAlaLeuLeuTyrArgValLeuThr 371
2493 CCACCGCGTACAAAACGGCAGTGTGACGCTTCGGCAACGCTAAGCA 2542
371 OGluProValIys.....LeuThrLeuAlaGlyGlyAlaGlnGly 384
2543 ACGTAGCATTCGCCACTCAACGGTAATGTCTCCCTAGCCGATAAGCA 2592
385 .....GlnGlyAspIleValAlaThr..... 391
2593 GATTTCCATTTGAAAAGCGCGCTTTACGGCAAAATCAGCGCGGCA 2642
392 .....GluLeuProProIleProGlyAlaSerSerGlyProIle 404
2643 GGATACGGATACACTTAAAGACAGCGAATGACGCTGCCGTCAGCA 2692
404 uAspAlaIleLeu...AlaSerGlnAlaIleArgTyrPhe...GlyAlaThrA 419
2693 CGGAATTAGCAATTTAAACCTTGACAACGCCACCATTAACCTCAATTC 2742
419 rGAlaValAspSerLeuSerIleAspAsnAlaThrTyrValMet..... 433
2743 GCCTATCGCCAGATCGGCGAGGGCGCAACCGGCACTGGCAGATGC 2792
434 .....ThrAsp..... 435
2793 GCCGCGCGCGCTTCCGCGCTTCGCGCGCTTCCTATTATTCGTTAC 2842
436 .....AsnSerAsnValGlyAlaLeuArgLeuAla 446
2843 CGCCAACTCGGTAGAA.....TCCGTTTCAAC 2871
446 erAspGlySerValAspPheGlnGlnProAlaGluAlaGlyAlaArgPhe 462
2872 ACGGTACGGTAACGGCAATTTGACGGTATGAGCAATTCGCTTTAT 2921
463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478
2922 GTCCGAACCTTTCGCTACCGCAGCAAAATTAAGCTGGCGCAAGTT 2971
478 nValPheAlaAspLeuGlyLeuSerAspIysLeuValValMetArgAspA 495
2972 CCGAAGCACTTACACTTGGCGGTCAACAATACGGCAAGCAACTGCA 3021
.....
495 IAserGlyGlnHisArgLeuLeuValArgAsnSerGlySerGluProAla 511
3022 ACCCTGCAACAATTTGACGCTAGTAGAGAAAGAAAGCAACAAACGCTGTC 3071
512 SerGlyAsnThrMetLeuLeuValGln.....ThrProArgIle 524
3072 CGAAAACCTTAATTTACCTTGCAAAACGAA.....CAGCTGATGCGG 3115
524 ySerAlaIleAlaThrPheThrLeuAlaIleAsnIysAspGlyValAspIleG 541
3116 GCGCGTGGCGGTATACCACTCATCCGCAAAAGACGGGAGTCCGCTGCAT 3165
541 IyThrTyrArgTyrArgLeu..... 547
3166 AATCGGTCAAGAAACAGAGCTTTCGCAAAACTCGCAAGCGAGAAC 3215
547 ..... 547
3216 CAAAAACAGCGCGCAAAAAGACACGCGCAAAACCTTGACGCGGTATG 3265
548 .....AlaIleAsnGlyAsnGlyGlnTyr.....SerLeuValG 559
3266 CGCGCGGCGCGGATGCGCTCGAAAAAGACAAACCGTTGCCGAACCGGC 3315
559 IyAlaIleLysAlaProProAlaPro.....LysPro 568
3316 CGCGCAGCGAGCGGGGAAATGTCGCGATTTATGACAGCGGAGAGAGAA 3365
568 OAlaProGlnProGlyPro..... 574
3366 AAAACGGTGCAGCGCGATTAAGACACCGCCTTGCGAAACAGCGGAA 3415
574 ..... 574
3416 CGAAACCGCGCGCGCTACACACGCTTCCCGCGCGCGCGCGCGCG 3465
575 ..GlnProGlyProGlnProProGlnProProGlnProProGlnPro 590
3466 CGGATTTGCGCGCAACTGCACCCCAACCGCAGCGCCACCGAGCGGA 3515
590 rGlnArgGlnProGluAlaProAlaProGlnProProAlaGlyArgIle 606
3516 CCTATACGCGGTATAGCAATAGCGGTGTGAGTGAATTTCCGCCACGC 3565
606 uLeu...SerAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 622
3566 TCAACAGCGTT.....TTCGCGGTACAGGACGAATTAACCGGTATT 3609
622 euAlaSerThrLeuTyrAlaGlnIleSerAsnAlaLeuSerIysArgLeu 638
3610 GCCGAACACCGCGCGAAC.....GCCGTTTGACAAAGCGGCAT 3647
639 GlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaIleArgIleArg 655
3648 CCGGAGACCCAAACACTACCGCTTCGCAAGATTTCCGCGCTACCGCAAC 3697
655 euAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArgPheAspGln 672
3698 AAACCGACCTCGCGCAAAATCGGTATCGCAAGAAAACCTC..... 3735
672 yValAlaGlyPheGlnLeuGlyAlaAspHisAlaValAlaValAlaGly 688
3736 GCGAGCGGCGCGCTCGCATCTCTTGGCAACCGCAACCGCAAAACAC 3785
689 GlyArgTyrHisLeuGlyGlyLeuAlaGlyTyrThrArgIleAspArgIle 705
3786 CTTT...GACGACGGCATCGGCAACTCGCAGCGCTTGGCCAGCGCGG 3832
705 yPheThrGlyAspGlyGlyGlyHisThrAspSerValHisValGlyGlyT 722
3833 TTTTGGCAATACGGCATCGACAGGTCTTACATCGGCTACAGCGCGGCG 3882
722 yAlaIleThrTyrIleAlaAsnSerGlyPheTyrLeuAspAlaThrLeuArg 738

```

```

3883 GCGGCTTTTACGACGGCAGCCTT.....TCAGACGGC..... 3915
||||| :||||| :|||||
739 AlsterAglendulansspherelysValAlaGlySerAspGlyTyrAl 755
3916 .ATGAGGCAAAATCCCGCCGCGCTGTCGATTACGGCATTACGGCAC 3964
::: ||||| :||| :||||| :|||
755 aVallySGlyLysTyrArgThr.....HisGlyValGlyAlas 768
3965 GATACCGCGCCGCT.....TTCGGCGATTTCGGCATCGAA 3999
||||| :||||| :|||||
768 erleuGluAlaGlyArgArgPheAlaHisAlaSerGlyTyrPheLeuGlu 784
4000 CCGCAGCATTCGGCGCAACCGCTATTTCGTCGCAAAAGCGGATTACCGCTA 4049
||||| :||| :||| :|||||
785 ProGlnAlaGluLeuAlaValPheArgValGlyGlySerTyrArgAl 801
4050 CGAAAC...GTCAATATGCGCAACCGCGCTTCGATTCAACCGCTACC 4096
||| :||| :|||
801 aAlaAsnGlyLeuArgValArgAspGlyGlySerValLeuGlyA 818
4097 GCGCGGCGATTAAAGCAGATTATTATTCAAACCGCGCAACACATTCC 4146
||| :||||| :||| :|||
818 rgleuGlyLeuGluValGlyLysArgGlyLeuAlaGlyGlyArgGln 834
4147 ATCAGCGCTATTGAGCCTGCTCTATACGATGCGCGCAAGT 4196
::: ||||| :||| :|||
835 ValGlnProTyrIleLysAlaSerValLeuGlnLeuPheAspGlyAlaG 851
4197 CCGAAGCAGCGCTCAATACCGCGCTATTGCTCAAGATTTCGCGCAAAACC 4246
::: ||||| :||| :|||
851 yThValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThra 868
4247 GCACTGGCGAATGGGCGTAACCGCGCAATCAAGTTTCAGCGCTCC 4296
||| ||||| |||
868 rg...AlaGluLeuGly.....LeuGly 874
4297 CTCGACGCTGCGCGCCAAAGCGCGCACTGGAAGCG.....CAACA 4340
::: ||||| :||||| :|||
875 MetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGlyT 891
4341 CAGCGCGGCGATCAAAATTAGCTACCGCTG 4371
||| ||| |||||
891 rSerLysGlyProLysLeuAlaMetProTyr 901

seq_name: p1r2:S15204

seq_documentation_block:
pertactin - Bordetella parapertussis
N:Alternate names: outer membrane protein P70
C:Species: Bordetella parapertussis
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 08-Oct-1999
C:Accession: S15204; S14659
R:IL, L.J.; Dougan, G.; Novotny, P.; Charles, I.G.
Mol. Microbiol. 5, 409-417, 1991
A:Title: P.70 pertactin, an outer membrane protein from Bordetella parapertussis: cloning
A:Reference number: S15204; MIMD:91251771
A:Accession: S15204
A:Molecule type: DNA
A:Residues: 1-922 <LIL>
A:Cross-references: EMBL:X54547; NID:g939761; PIDD:CA38419.1; PID:g939762
C:Genetics:
A:Gene: prn
C:Keywords: membrane protein

```

```

alignment_scores:
  Quality: 295.00      Length: 1083
  Ratio: 0.602        Gaps: 52
  Percent Similarity: 45.245      Percent Identity: 22.253

```

```

alignment_block:
US-09-303-518D-649 x S15204

```

```

Align seg 1/1 to: S15204 from: 1 to: 922

1417 GGGGAAACCAAGGCTGATCAGCGTGGCGACGGTACAGTATTGGA 1466
||||| :||| :|||
45 GlyIuArgGlnHisGlyIleHisIle..... 53
1467 TCAGCAGCAGCAGTAAAGCAAAACAGCCTTAGTGAATCGGCT 1516
::: ||||| :||| :|||
54 LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrIleI 70
1517 TGGTCAGCGGAGGCGTACGGTGCACACTGAATGCCGATATCAGTTAAC 1566
||||| :||| :|||
70 yValSerGlyArgGlnAlaGlnGlyValLeuLeuGlnAsn..... 83
1567 CCCGCAACACTATTTGCGGCTTCGGGGGAGCGAGTTGATTAAACG 1616
::: ||||| :||| :|||
84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerG 99
1617 GCATTGCGCTTCGTCACCGCTATTCAAAATACGATGAGGGCGCATGA 1666
| ||| |||||
99 y.....GlnLeuPheAspGlyVal.... 106
1667 TTGTCACACCATCAAGCAAAAGATCCACCGTACCATTAAGGCAAT 1716
::: ||||| :||| :|||
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGly 118
1717 AAA.....GATTATGCTACACCGGCAATTAACACAGCTTGATAG 1757
||| ||||| :|||
119 LysLeuValAlaAspHisAlaThrLeuAlaAsn..... 129
1758 CAAAAAGAAATGCTTACACCGTTGGTTGGCGCAAAATACGACCA 1807
::: |||||
130 .....ValSerAspThrArgA 135
1808 AAACGAGCGCGGCTCAACCTGTT.....TACAGCGCGCC 1845
::: |||||
135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluAlaGlnAlaSer 151
1846 GCAGAGACCGGACCGCTGCTGCTTCGCGGAGCAAAATTTA.....AA 1889
||| ||||| :|||
152 IleAlaAspSerThrLeuGlnGlyAlaGlyValArgValGluArgG 168
1890 CCGCAACATCAGCAAAACAGCGCAACTGTTTTCACGGC..... 1932
::: ||||| :||| :|||
168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHis 185
1933 .....AGACCAACACCGCGCTTACATCATTTAAACGACAT 1971
|||
185 leGlyThrLeuGlnProLeu..... 191
1972 TGGTCGCAAAAGAGGCGATTCCTCGCGGAGAAATCGTGGGACACGA 2021
||| ||| :||| :|||
192 .....GlnProGlnsPneuProPserArgValValLeuGlyAspPh 206
2022 CTGATCAACCGC.....ACAT 2038
:::
206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheVal 223
2039 TTAAGCGGAAACCTTCAATTAAGCGGAGCGCGGCTGTTCCCGC 2088
||| :||| :||| :|||
223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239
2089 AATGTGCGCAAAAGTGAAGGCGATTGGCATTTGACAAATCAGCCCAACG 2138
:::
240 AlaAlaGly.....ValAlaAl 245
2139 AGTTTGGTGTGCGACCGCATCAAGGCCACACATCTGATACGCTCGG 2188
||| :||| :||| :|||
245 ameLAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262
2189 AC..... 2190

```

```

262 spalaProAlaGlyGlyAlaValaProGlyAlaValaProGlyGlyAla 278
2191 .....TGACGGGTGACAA 2207
279 ValProGlyGlyPheGlyProLeuLeuAspGlyTrpGlyAla..... 293
2208 TTGTGTGAAAAAACCATACCGACGATTAAGTATGCTTCATGTGACTA 2257
293 ..... 293
2258 AGACCGCATCAGC...GCCATGTGATCTTCCGATCAGCTCATTTA 2304
294 .....AspValSerAspSerThrValAspLeuAlaGln..... 304
2305 AATCTCACAGGCTTGCCACACTCAACGGCAATCTTAGTCAATAGCCGA 2354
305 SerIleValGluAlaProGlnLeuGlyAlaAlaIleArgAlaIleArgG 321
2355 TCACAGCTTATACAGTCAGCCACAAC.....GCCACCCAAACGGCACG 2398
321 yAlaArgValThrValSerGlySerLeuSerAlaProHisGlyAsnV 338
2399 TTAGCCCTGGGGCAATGCCCAAGCAACATTTAAT..... 2433
338 aIleGlnThrGlyGlyAlaAlaArgArgPheProProAlaSerPro 354
2434 ...CAAGCCACATTTAAACGGCAACACATCGCTTCGGCAATGCT...TC 2477
355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaLeu 371
2478 ATTATATCTTAAGCAGCAGCCGCTACAAAACGGCACTGACGCTTCGG 2527
371 uTyArgValLeuProGlnProValys.....LeuThrLeuAlaG 385
2528 GCACCCCTAAGGCAACGTAAAGCATTCGCACTCAAGGTAATGCTCC 2577
385 GlyAlaGlnGly.....GlnIleAspIleVal 394
2578 CTAGCCGATAAGGCACTATTCATTTGAAGCAGCCGCTTACCGGACA 2627
395 AlaThr.....GluLeuProProIleProGlyAla 404
2628 AATCAGCGCGCGGAGATAGCGCATTTACCTTAAGACAGCGAATGGA 2677
404 AserSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTrp 420
2678 CCCTGCCGTGAGCAGCGAATTAGCGAATTTAACTTGACAAAGCCGAC 2727
420 hr...GlyAlaThrArgAlaValaAspSerLeuSerIleAspAsnAlaThr 435
2728 ATTACACTCAATTCGGCCCTATGCGCACGATGCGGAGGCGCAACCGG 2777
436 TrpValMet..... 438
2778 CAGTCGACAGATGCGCGCGCGCGCTTCGGCGCTTCGGCGCTGCC 2827
439 .....ThrAsp.....AsnSerAsnValGlyA 446
2828 TATTTATCCGTTACACCCGCAACTTCGGTAGAA..... 2859
446 IAlaLeuArgLeuAlaSerAspGlySerValaAspPheGlnGlnProAlaGlu 462
2860 ...TCCGCTTCAACAGCTGCGGTAAAGCGCAATTAAGACGCTGACGG 2906
463 AlaGlyArgPheLeuValLeuMetValAsp...ThrLeuAlaGlySerG 478
2907 AACATTCGCTTTATGTCGAACCTTCGCTACCGCAGCAGCAAAATGA 2956
478 yLeuPheArgMetAsnValaPheAlaAspLeuGlyLeuSerAspIleuV 495
2957 AGCTGGCGGAAGTTCCGAAGCACTTACACTTCGCGGTACACAAATACC 3006
495 aIvalMetArgAspAlaSerGlyGlnHisArgLeuTrpValaArgAsnSer 511
3007 GGCACAGCAACTTCGAAGCCTCGAACATTTAGCTAGTAGGAAGAAAGA 3056
512 GlySerGlnProAlaSerGlyAsnThrMetLeuLeuValGln..... 525
3057 CAACAAACCGCTGTCGAAACCTTAATTTACCCCTGCAAAACGAA.... 3102
526 ...ThrProArgGlySerAlaIleAlaThrPheThrLeuAlaAsnIysAspG 541
3103 ...CAGCTGATGCGCGGCTGTACCAACTCATCTCCGCAAGACGGC 3150
541 yLysValaAspIleGlyThrTyArgTyArgLeu..... 552
3151 GAATTCCGCTTCATTAATCCGTCAAAGACAGAGCTTTCGCAAACT 3200
552 ..... 552
3201 CGGCAAGCGCAGAGCCAAACAAAGCGGAAACAGACCGGCAAGCC 3250
553 .....AlaAlaAsnGlyAsnGlyGlnTrp. 560
3251 TTGACGCGCTGATTCGCGCGCGCGGATTCGCGCAAAAGACAGAAAGC 3300
561 ...SerLeuValGlyAla..LysAlaProProAla.....ProLysPr 573
3301 GTTCCCGAACCAGCGCGCGCAGGCAAGCGGGAATGTCGCAATTATGCA 3350
573 oAlaPro...GlnProGlyProGlnProGln..... 584
3351 GCGGAGAGAGAGAAACAGGCTGCAAGCGGATTAAGACACCCGCTTGG 3400
585 .....ProPro... 586
3401 CGAAACAGCGGAAGCGGAACCGCGCGGCTACACCGCTTCGCGCGC 3450
587 .....GlnProProGln..ProProG 593
3451 GCCCGCGCGCGCGCGGATTTCCGCACTGCAACCCCA..... 3492
593 InProProGlnProPro..GlnProProGlnArgInProGlnAlaProAl 609
3493 ...CCGACGCGCCACCGCAGCGCAGCTGATCAGCGCTTATGCGCAATGCG 3541
609 aProGlnProProAlaGlyAlaArgGlyLeu...SerAlaAlaAlaAsnAla 625
3542 GTTGAAGTGAATTTCCGCGCAGCTCAACAGCTT.....TTCGCGCTA 3585
625 IAlaAlaAsnThrGlyAlaGlyAlaGlyLeuAlaSerThrLeuTrpTyAlaGlu 641
3586 CAGACGCAATTTAGACCGCTATTTGCCGAGACCGCGCCGCAAC..... 3627
642 SerAsnAlaLeuSerIleArgLeuGlyGlyLeuArgGlnGlnLeuAspAsnAl 658
3628 ...GCCGTTTGGACAAGCGCATCCGCGACASACAAACATACCTGTCG 3673
658 aGlyGlyAlaTrpGlyArgGlyLeuPheAlaGlnArgGlnLeuAspAsnAl 675
3674 AAGATTTCCGCGCGCTACCGGCAACAAACGACCTGCGGCAATCGGTATG 3723
675 rGlyAlaGlyArgArgPheAspGlnLysValaIleGlyPheGlyLeuAla 691
3724 CAGAAAAACCTC.....GGCAGCGCGCGCGTCCGCACTGCTT 3761
692 AspHisAlaValaIleAlaValaIleGlyAlaGlyArgTrpHisLeuGlyLeuAl 708
3762 TTGCGCAACACCGGCAAGCAACCTTC...GACGACGCGATCGGCACT 3808
708 aGlyTyTrpArgGlyAspArgGlyPheThrGlyAspGlyGlyGlnHisT 725
3809 CCGCAGCGCTTGGCCACGCGCGCGCTTTCGGGCAATACGCGCATCGACAG 3858
725 hnsPheArgValHisValGlyGlyTyArgAlaThrTyTrpIleAlaAsnSerGly 741

```


[illegible]

```

2905 GGAACATTCGGCTTATGTGSACTCTTCGC.....TACCGACGCA 2948
|||||: : : |||||: |||
823 G|T|H|L|e|u|e|u|a|s|p|e|r|g|l|e|u|a|s|n|g|l|y|a|s|p|a|s|e|r|v|a|l|S|e|r|a| 839
2949 CAATGTGAGCTGGCGGAGAAAGTTCCGAAAGGCACTTACACCTTGCGGCGTCA 2998
|||||: : : |||||: |||
839 p|G|n|L|e|u|a|l|e|u|a|s|n|g|l|y|a|s|n|h|a|l|a|g|l|y|a|s|n|h|t|H|v|a|l|a| 856
|||||: : : |||||: |||
2999 ACNAT...ACCGGC...AAGCAACTGCAGGCTCGAACATATGACGTA 3042
|||||: : : |||||: |||
856 s|n|s|e|r|l|e|t|H|g|l|e|g|l|y|l|u|p|r|o|t|H|s|e|r|t|H|g|l|y|l|e|u|s|v|a|l| 872
3043 GTGGAAGGAAAGACAAACACCGCTGTCGAAAACCTTAATTTGACCGT 3092
|||||: : : |||||: |||
873 A|s|p|H|e|a|l|a|l|a|s|p|r|o|t|H|g|l|n|p|H|e|g|l|n|a|s|n|a|l|a|g|l|n|p|H|e|s|e|r|l|e 889
3093 GCAAAACGAA...CACGTGATCCGCGCGCTGCGCTTACCAACTCATCC 3139
|||||: : : |||||: |||
889 u|a|l|a|g|l|y|s|e|r|g|l|y|t|y|r|a|l|a|s|n|e|t|g|l|a|t|y|r|a|s|p|l|y|t|H|l|e|u|a|l|g 906
3140 GCAAGACGGCGAGTTCGCCCTGCATTAATCCGCTCAAGACAGACTT 3189
|||||: : : |||||: |||
906 l|u|s|p|a|s|n|a|s|p|l|y|t|y|r|l|e|u|..... 913
3190 TCCGACAAACTCGGCAAGGACGAGCAAAACAGCGGAGAAAGACAA 3239
|||||: : : |||||: |||
913 ..... 913
3240 CGCGCAAAAGCTTGACGCGCTGATGCGCGCGCGCGATGCCGTGAAA 3289
|||||: : : |||||: |||
914 .....A 914
3290 AGACAGAAACGTTGCCGAACCGCGCGCGAGCGGCGGAGAAATGTC 3339
|||||: : : |||||: |||
914 r|g|s|e|r|g|l|n|g|l|u|a|l|t|H|p|r|o|..... 921
3340 GGCATTATGACGGGAGAGAGAAAAAGCGGTGACAGCGGATTAAGA 3389
|||||: : : |||||: |||
921 ..... 921
3390 CACCGCTTGCGCAACAGCGGAGCGGAAACCGCGCGGTCACACCG 3439
|||||: : : |||||: |||
922 .....S|e|r|P|r|o|t|a|s|p|r|o|t|a|s|p|r|o|t|H|p|r 930
3440 CCTTCCCGCGCGCGCGCGCGCGGATTTGCGCAACTGCAACCC 3489
|||||: : : |||||: |||
930 r|o|a|s|p|r|o|t|a|s|p|r|o|t|H|n|a|s|p|r|o|t|a|s|p|r|o|t|H|p|r|o|g|l|u|p|r|o 946
3490 CAACGAGCGCCCAACCGGACGCGGACCTGATCAAGCGGTATGCAATAG 3539
|||||: : : |||||: |||
947 T|H|p|r|o|l|a|t|y|r|g|l|n|p|r|o|v|a|l|l|e|u|a|s|n|..... 955
3540 CGGTTGAGTAATTTCCCGCAGCGCAACAGCGTTTCCCGCTACAGG 3589
|||||: : : |||||: |||
956 .....A|l|a|y|s|v|a|l|g|l|y|t|y|r|l|e|u|a|s|n|l|e|u|a|g|l|a|l|a|..... 968
3590 ACGAATGACCGCGTATTTGCGAGACCGCGCAACCGCGGTTGGACA 3639
|||||: : : |||||: |||
969 .....A|s|n|g|l|n|a|l|a|p|H|e|t|e|t|g|l|u|a|t|g|a|r|a|s|n|l|a|l|a|g|l|y| 982
3640 AGCGGC.....ATCCGGACACCAACACTACCG 3688
|||||: : : |||||: |||
983 A|s|p|g|l|y|n|t|H|l|e|u|a|s|n|l|e|u|a|g|l|l|e|g|l|y|a|s|p|l|y|t|H|y|t|H 999
3669 TTGCGAATTTCCCGCGCTACCGCAACAGCGAGCTGCGCAATCG 3718
|||||: : : |||||: |||
999 T|A|l|a|l|a|g|l|y|n|l|e|u|a|l|a|.....G|l|n|h|s|g|l|n|a|s|p|H|s|e|r|t|H|v|a|l|g 1014
3719 GTATGCAAAAACCTCGGACGCGCGCG..... 3747
|||||: : : |||||: |||
1014 I|n|t|e|r|s|e|r|g|l|y|a|s|p|l|e|u|h|e|r|g|y|a|t|g|t|y|r|g|l|y|t|H|a|s|p|l|y|g|l|u|t|p|r 1030
3748 ..GTGGCATCTG.....TTTTCGACAAACGCGGACCGAAACACTT 3788

```

```

1031 M|e|t|H|e|u|g|l|y|l|e|v|a|l|g|l|y|t|y|r|s|e|r|a|s|p|a|s|n|g|l|n|g|l|y|s|p|e|r|a|y|s|e 1047
|||||: : : |||||: |||
3789 CGAGACGGCATCGGCAACTCGGACGCGCTTGGCCACGCGCGGCTTTTCG 3838
|||||: : : |||||: |||
1047 r|s|e|r|e|t|H|g|l|y|t|H|a|r|g|a|l|a|s|p|a|s|n|g|l|n|h|s|g|l|y|t|y|r|a|l|a|l|g 1064
|||||: : : |||||: |||
3839 G|G|.....CAATCGCATTCGACAGGTTCTCAATCGCG 3870
|||||: : : |||||: |||
1064 l|y|l|e|u|t|H|s|e|r|t|H|p|H|e|g|l|n|h|s|l|y|s|g|l|n|y|s|..... 1076
3871 A|T|C|A|G|C|G|G|G|C|G|G|.....G|T|T|T|A|G|C|A|G 3896
|||||: : : |||||: |||
1077 .....G|n|c|l|y|a|l|a|t|y|r|l|e|u|a|s|p|a|s|n|t|H|p|l|e|u|g|l|n|t|y|r|a|l|a|t|H|p|H|e|s|e 1091
3897 CGGACGCTTTCAGACGCGGATCGAGCAAAATCGCGCGCGCTGTCG 3946
|||||: : : |||||: |||
1091 r|a|s|n|a|s|p|a|s|e|r|g|l|n|h|s|g|l|u|a|s|p|l|y|.....V|a|l|a|s|p|H 1103
3947 A|T|T|A|C|.....G|G|C|A|T|T|C|A|G|C|A|G|A|T|C|G|C|G|C|G|T|T|C|..... 3981
|||||: : : |||||: |||
1103 l|s|t|y|r|n|h|s|e|r|s|e|r|c|l|y|l|e|t|H|e|a|l|a|s|e|r|l|e|u|g|l|n|a|g|l|y|t|y|r|g|l|n|t|p|r 1119
|||||: : : |||||: |||
3982 .....G|C|G|G|A|T|T|C|G|G|C|A|T|G|A|C|C|G|C|A|T|C|G|G|C|G|C|A|G|G|C|G|T|A 4022
|||||: : : |||||: |||
1120 l|e|u|p|r|o|g|l|y|a|r|g|l|y|a|l|l|e|g|l|n|p|r|o|g|l|n|a|g|l|n|l|e|t|y|r|g|l 1136
|||||: : : |||||: |||
4023 T|T|T|G|T|C|C|A|A|A|A|G|G|G|A|T|T|A|C|C|G|T|A|C|G|A|A|A|C|..... 4056
|||||: : : |||||: |||
1136 n|g|l|y|a|l|n|c|l|n|a|s|p|a|s|p|H|e|r|t|H|a|l|a|s|n|a|l|a|r|g|a|l|a|r|g|a|l|s|e|r|g 1153
|||||: : : |||||: |||
4057 .....G|T|C|A|A|T|A|T|C|G|C|A|C|C|C|C|G|G|C|T|T|G|C|A|T|T|A|C|C|G|C|T|A|C 4095
|||||: : : |||||: |||
1153 I|n|s|e|r|g|l|n|g|l|y|a|s|p|l|l|e|g|l|n|t|H|p|r|o|..... 1161
4096 CGCGGCGCATTAAGGAGATTAATTCATTAACCGCGCAACACATTC 4145
|||||: : : |||||: |||
1162 A|r|g|l|e|u|g|l|y|l|e|u|n|h|s|e|r|g|l|u|t|p|r|.....A|r|g|t|H|a|l|a|V|a|l|H|s|..... 1174
|||||: : : |||||: |||
4146 C|A|T|C|A|G|C|G|T|A|T|T|G|A|C|C|T|G|T|C|.....T|A|T|A|C|G|A|C|C|C|T|T|G|G|G|C|A 4192
|||||: : : |||||: |||
1175 V|a|l|l|e|r|p|r|o|t|H|l|e|u|a|s|p|l|e|u|a|s|n|t|y|r|t|H|n|a|s|p|r|o|t|H|s|e|r|t|H|g 1191
|||||: : : |||||: |||
4193 A|A|G|T|C|G|A|A|C|G|G|C|T|C|A|A|T|A|C|C|G|G|T|A|T|G|G|C|T|A|G|A|T|T|G|C|G|A|A 4242
|||||: : : |||||: |||
1191 l|u|l|e|g|l|n|u|a|s|p|a|l|a|s|e|r|t|H|p|r|.....l|l|e|r|a|s|p|a|l|a|l|a|l|y|s 1205
|||||: : : |||||: |||
4243 A|C|C|C|G|C|.....A|G|T|C|G|G|A|T|G|G|G|C|T|A|A|C|C|G|G|A|A|T|C|A|A|G|T|T|C|A|C 4289
|||||: : : |||||: |||
1206 G|l|n|a|r|g|l|y|g|l|l|e|l|y|s|v|a|l|g|l|y|a|l|t|H|g|l|y|a|s|n|l|l|e|r|g|l|n|.....A|r 1221
|||||: : : |||||: |||
4290 G|C|T|G|C|C|T|C|A|C|G|T|G|C|G|C|G|C|C|.....A|A|G|C|C|G|C|A|C|T|G|G|A|G 4333
|||||: : : |||||: |||
1221 g|v|a|l|s|e|r|l|e|u|a|r|g|l|y|s|e|r|v|a|l|a|l|a|t|p|l|n|g|l|y|s|e|r|a|s|p|a|s|p|H|e|a 1238
|||||: : : |||||: |||
4334 C|G|C|A|C|A|C|A|G|C|G|G|C|G|C|A|A|T|A|G|C|T|A|C|C|G|T|G 4371
|||||: : : |||||: |||
1238 l|a|g|l|n|t|H|a|l|a|g|l|y|p|H|e|u|s|e|r|e|t|H|v|a|l|l|y|t|p|r 1250

```

seq_name: pir2:F85862

seq_documentation_block:

hypothetical protein yfal, [imported] - Escherichia coli (strain O157:H7, substrain ED C:Species: Escherichia coli C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001 C:Accession: F85862 R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May iler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamovits, K.; Apoda Nature 409, 529-533, 2001 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7. A:Reference number: AB5460; M01D:21074935; PMID:11206551 A:Accession: F85862 A>Status: preliminary


```
1954 .....ATCATTTAAAGCACCATTGGTCGCAAAA 1983
531 aValaspLeuValaglyGlyAsnAspMetHisIleglyIysaspGlyLys 547
1984 GAGGCG.....ATTCCTCCGGGGGA.....ATCGT 2009
548 AspGlyValIytrValIvalIleaspAlaglyaspGlyGlnValSerIleuVal 564
2010 GTGGGACAAACGACTGGATCAACCGCACATTTTAAAGCGGAAACTTCGCAA 2059
564 aAsnAspAsnGlnIytrLeuGlyThr.....GlnI 575
2060 TTTAAAGCGGACAGCGCGTGTTCGCCGCAATGTTCGCAAGGAAAGCG 2109
575 leaIleSerGlyThrLeuMetValSerAspAsn.....SerGlnLeuGly 589
2110 GATGGCATTTGAGCAATCAAGCCGCAAGCAGTTTGGTGTGCGACCGCA 2159
590 TyrThrIstIytrAsnArg.....GlnValIlePheThrAspIysProG 604
2160 TCAAAAGC.....CACACAACTGTACACGTTGCGACTGGA 2194
604 nGluSerValMetGluIleThrAlaAsnValAspThrArgSerThrThrT 621
2195 CGGGTGTGACAAATTTGTGCAAAAAACCAATTCCGACGATTAAGTGAT 2244
621 hrGlnHisGlyArgAspIleGln.....MetArgAlaAspGlyGlnVal 635
2245 GCTTCATTGACTAAGACCGACATCGCGGCAATGTCGATCTTGGCGATCA 2294
636 AlaValaspAlaglyValaspThrGlnIytrpGlyAlaLeuMetAlaasp 652
2295 C.....GCTCATTTAAATCTCACAGGGCGTTG 2320
652 rSerGlyGlnHisGlnaspGlnGlySerThrLeuThrIystrGlyAlaG 669
2321 CCACACTCAAGCGCAATCTTAGTCAAAATGCGATACACGTTATACATC 2370
669 IythrLeu.....GluLeuThrAlaSerGlyThrGlnIleSerAlaVal 683
2371 AGCCCAACAGCCACCCAAACGCAACCTTAGCCTCGTGGCAATGCCCA 2420
684 .....ArgValGlnGlnGlyThrLeuGln.....GlyAspValAla 695
2421 AGCAACATTTATCAAGCCACATTA.....AAGCGCAACACAT 2458
695 aspIlePheProIytrAlaSerSerLeuTrpValGlyaspGlyAlaThrP 712
2459 CGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACGACCGGTACAAAC 2508
712 heValThrGly.....AlaaspGlnaspIleGlnSer 722
2509 GGCAGCTGAGCGCTTCGCGGACGCTAAGCAACGTAAAGCATTCGCG 2558
723 IleaspAlaThrSerSer..... 728
2559 ACTCAACGGTATGTCCTCCATAGCCGATAAAGCAGATATTCATTTTGA 2608
729 .....GlyThrIleaspIleSerAspGlyThrValLeu..... 739
2609 GCAGCGCGCTTACCGGACAAATCAAGCGGCGGCAAGATAGCGCATTAAC 2658
740 .....ArgLeuThrGlyGln.....AspThrSerValAla 749
2659 TTTAAAGACAGC.....GAATGACGCTGCGCGTCCGAGGAGGA 2696
750 LeuAsnAlaSerLeuPheAsnGlyaspGlyThrLeuValAsnAlaThrAs 766
2697 A.....TTAGCAATTTAAAC..... 2712
766 pGlyValThrLeuThrGlyGlnLeuAsnThrIystrLeuGluThrAspSer 783
2713 .....CTTGACAAAGCCACCATTTACATCAATTCGCGCTATCGCCAC 2754
```

```
783 eutThrIyLeuSerAsnValThrValAsnGlyAsnLeuThrAsnThrSer 799
2755 GATGGCGGACGGGCGCAAAACCGGACGTCCGACACATCGCGCGCGCCG 2804
800 GlyAlaValSerLeuGlnAsnGlyValAlaGlyasp..... 811
2805 TTCGGCGCGTTCGGCGCGTCCCTATTATTCGTTACACCGCCAACTTCG 2854
811 ..... 811
2855 TAGAATCCCGTTTCAACACGCTGACGTAACGGCAATTAACGCTGAC 2904
812 .....ThrLeuThrValAsnGlyAspTyrThrGlyGly 822
2905 GGAACATTCGCGCTTATGTCCGAACCTTCGCGC.....TACCGAGGGA 2948
823 GlyThrLeuLeuAspSerLeuLeuAsnGlyaspAspSerValSerAs 839
2949 CAATTTGAAGCTGGCGGAAAGTTCCGAAGCAGCTTACACCTTGCGGTCA 2998
839 pGlnLeuValMetAsnGlyAsnThrAlaGly***ThrThrValValAla 856
2999 ACAAT...ACCGGC...AACGACCTGCAAGCCTCGACAAATTAACGCTA 3042
856 snSerIleThrGlyIleGlyGlnProThrSerThrGlyIleIystrVal 872
3043 GTGGAAAGAAAGACAAACAAACCGCTGTCGAAACCTTAATTCACCC 3092
873 AspPheAlaIleAspProThrGlnIlePheGlnAsnAlaGlnPheSerIe 889
3093 GCAAAACGAA...CACGTGATGTCGCGCGCGGTGCGTTACCACTCATCC 3139
889 uAlaGlySerGlyTyrValAsnMetGlyAlaTyrAspTyrThrLeuVal 906
3140 GCAAAACGCGGAGTTCGCGCTGCATTAATCCGCTCAAAAGAACAGAGCTT 3189
906 IuaspAsnAsnAspTyrIyLeu..... 913
3190 TCCGACAAACTGCGCAAGCAGAACGCAAAAAACAGCGCGGAAAAAGACA 3239
913 ..... 913
3240 CGCGCAAAAGCTTGACCGCTGATTCGGCGCGGCGCGATGCGTCGAA 3289
914 .....A 914
3290 AGACAGAAAGCTTGCCGAACCGCGCGCGAGCAGCGCGGAGGAAAAATGTC 3339
914 rSerGlnGlnValThrProPro..... 921
3340 GGCATTATGCGAGCGGAGAGAAAGAAAAACGGGTCCAGCGGATAAAGA 3389
921 ..... 921
3390 CACCGCGTTGGCGAAACAGCGGAGGGAACCGCGCGGTACACCGG 3439
922 .....SerProProAspProAspProThrP 930
3440 CCTTCCCGCGCGCGCGCGCGCGCGCGGAGTTTGGCGCAATGCAACCC 3489
930 roAspProAspProThrGlnAspProAspProThrProAspProGlnPro 946
3490 CAACCGGACCCCAACCGCAGCGGACCTGATCAAGCCGTTATGCAATAG 3539
947 ThrProAlaTyrGlnProValLeuAsn..... 955
3540 CGGTTTACGATGATTTCCGCGCAAGCTCAACAGGTTTTCGCGGTACAG 3589
956 .....AlaIystrValGlyGlyTyrLeuAsnThrLeuAlaAla... 968
3590 ACGAATTAGACCGCGTATTTGCGGAAGACCGCGCAACGCGTTTGACA 3639
```

```

969 .....AsnGlnAlaPheMetMetGluArgArgAspHisAlaGly 982
3640 AGCGGC.....ATCGGGACACCAACACATCCG 3668
983 AspGlyGlnThrLeuAsnLeuArgValIleGlyGlyAspTyrHisTyr 999
3669 TTGGCAAGATTTCGGCGCTACGCCAACAAACGACCTGGCCAAATTC 3718
999 rAlaAlaGlyGlnLeuAla.....GlnHisGluAspThrSerThrValG 1014
3719 GTATGCGAANAACCTCGGACGGGGCGC..... 3747
1014 LndSerGlyAspLeuPheSerGlyArgTrpGlyThrAspGlyGluTrp 1030
3748 ...GTGCGATCCCTG.....TTTGCACACCGGACCGCAAAACACCTT 3788
1031 MetLeuGlyIleValGlyGlyTyrSerArgAsnGlnGlyAspSerArgSe 1047
3789 CGACGACGCGATCGGCACTCGGACGGCTGCCACGGCGCGTTTTCG 3838
1047 rSerMetThrGlyThrArgAlaAspAsnGlnAsnHisGlyTyrAlaValG 1064
3839 GG.....CAATACGCATCGACAGGTTTCAATCGGC 3870
1064 LyleuThrSerSerTrpPheGlnHisGlyGlyGlnLys..... 1076
3871 ATCAGCGCGGCGCG.....GTTTTCACAG 3896
1077 .....GlnGlyAlaTrpLeuAspAsnTrpLeuGlnTyrAlaTrpPheSe 1091
3897 CGGACGCTTTTCAGACGCGCATCGAGGCAAAATCCGCGCGCGCTTCGC 3946
1091 rAsnAspValSerGlnHisGluAspGly.....ValAspH 1103
3947 ATTAC.....GGCATTCAGCATACCGCGCGCTTC..... 3981
1103 IeTyrHisSerSerGlyIleIleAlaSerLeuGlnAlaGlyTyrGlnTrp 1119
3982 .....GGCGATTTCGCGATCGACCGCATCGCGCGCAACCGCGTA 4022
1120 LeuProGlyArgGlyValIleGlnProGlnAlaGlnValIleTyrGln 1136
4023 TTTCGTCGCAAAAGCGGATTACCGCTACGAAAC..... 4056
1136 nGlyValGlnGlnAspAspPheThrAlaAlaAsnArgAlaArgValSerG 1153
4057 .....GTCAATATCGCCACCGCGCGCTTCATTCACCGCTAC 4095
1153 LndSerGlnGlyAspAspIleGlnTrp..... 1161
4096 CGCGCGGCGCATTAAGCGACATTAATTCATCAACCGCGCAACATTC 4145
1162 ArgLeuGlyLeuHisSerGlnTrp.....ArgThrAlaValHis..... 1174
4146 CATCACGCGCTATTGAGCTGTGC...TATACGATCGCGCTTCGGCA 4192
1175 ...ValIleProThrLeuAspLeuAsnTyrTyrHisAspProHisSerThr 1191
4193 AAGTCGCAACACGCGCTCAATACCGCGCTATTCGCTCAGATTCGGCA 4242
1191 LndLeuGlnGluAspAlaSerThr.....LndSerAspAspAlaValLys 1205
4243 ACCGCG...AGTCGGAATGGGGGTAAACCGCAATCAAGGTTTCAC 4289
1206 GlnArgGlyGlnIleLysValGlyValThrGlyAsnLndSerGln...Ar 1221
4290 GCTGTCCCTCCACGCTCGCGCGC...AAGCGCGCGCAACTGGAG 4333
1221 gValSerLeuArgGlySerValAlaIleTrpGlnLysGlySerAspSpha 1238
4334 CGGACACACGCGCGCATCAATTAAGCTACGCTGG 4371
1238 LndGlnThrAlaGlyPheLndSerMetThrValLysTrp 1250

```

```

seq_name: p1r2:A32560
seq_documentation_block:
outer membrane protein P.69 precursor - Bordetella pertussis
C:Species: Bordetella pertussis
C>Date: 07-Sep-1990 #sequence-revision 07-Sep-1990 #text-change 08-Oct-1999
C:Accession: A32560
R:Charles, I.G.; Dougan, G.; Pickard, D.; Chatfield, S.; Smith, M.; Novotny, P.; Mori
proc. Natl. Acad. Sci. U.S.A. 86, 3554-3558, 1989
A:Title: Molecular cloning and characterization of protective outer membrane protein
A:Reference number: A32560; MIMD:89264462
A:Accession: A32560
A:Molecule type: DNA
A:Residues: 1-910 <CHAS>
A:Cross-references: GB:J04560; NID:g144053; PIDN:AAA22980.1; PID:g144054
A>Note: It is uncertain whether Met-1 or Met-3 is the initiator
C:Keywords: membrane protein
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-910/Product: outer membrane protein P6.9 #status predicted <MAT>

```

```

alignment_scores:
Quality: 283.00 Length: 1100
Ratio: 0.581 Gaps: 52
Percent Similarity: 44.273 Percent Identity: 21.909

```

```

alignment_block:
US-09-303-518D-649 x A32560

```

```

Align seg 1/1 to: A32560 from: 1 to: 910

```

```

1366 AACAGCGCGCTGTCCAAATCGCAAAAGCGACCGCTGCACGTTCAAGCCAA 1415
||||| : : : ||| : : : : : : : : : : : : : : : : : : : : : : : :
38 AsnGlnSerIleValIleThrGlyGluArgGlnHisGlyIleHisIleGln 54
1416 AGGGAAACCAAGCGCTCGATCAGCGCTGGCGGCGACGATGATTTGG 1465
||||| : : : : : : : : : : : : : : : : : : : : : : : :
54 nGlySerAspProGlyValIleArgThrAlaSerGlyThrThrIle.... 69
1466 ATCAGACGCGACGATTAAGCGCAAAACAGCGCTTACTGAATCGGC 1515
69 ..... 69
1516 TTGGTCAGCGCGGCGGTACGCTGCACATGATCGCATATGAGTTCAA 1565
||||| : : : : : : : : : : : : : : : : : : : : : : : :
70 LysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGlnLys..... 83
1566 CCCGACAAACTCTATTTCGCGCTTCGCGGCGGACGTTTGATTTAAGC 1615
84 ....ProAlaIleGlnLeuGlnPheArgAsnGlySerValThrSerSer 99
1616 GCGATTCGCTTTCGCTTCACCGCTATTCAAATACGATGAAAGGCGGATG 1665
||| : : : : : : : : : : : : : : : : : : : : : : : :
99 Lys.....GlnLndSerAspGlyLeu.... 106
1666 ATTGTCAACCAATCAAGACAAAGAATCCACCGCTTACCATTAACAGCAA 1715
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGln 118
1716 TAA.....GATATTGCTTCAACCGCGCATTAACACAGCTTGATG 1756
||||| : : : ||| : : : : : : : : : : : : : : : : : : : : : : : :
118 LysLeuValIleAlaAspHisAlaThrLeuAlaAsnValGly..... 131
1757 GCAGAAAGAAATTCGCTACACGCGTTGGTGGCGAGAAAGATACGACC 1806
132 .....AspThrTrp 134
1807 AAACGACGCGCGGCTCAACCTGTT.....TACGACCGCGC 1844
||||| : : : ||| : : : : : : : : : : : : : : : : : : : : : : : :
135 AspAspAspGlyIleAlaLeuTyrValAlaIleGlyGlnAlaGlnAlaSe 151
1845 CGCAGACGCGACCGCTGCTTCGCGCGGAAACAAATTTA.....A 1888

```



```

612 GCGATCTGATGAGATGAGCCCAATAACCGGCAAGTCTCATCATATTG 661
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
284 yLysleuSerAlaSpSerValSerLysAspLysSerGlyAsnIleValI 301
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
662 CAAGTGGTATTCCTGG.....CTCGTGGTCGCAATACCTTTGGCAAA 705
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 euSerAlaLysGluGlyLualaGluIleGlyValIleSerAlaGln 317
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
706 AAT..... 708
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318 AsnGlnGlnAlaLysGlyLysLysLeuMetIleThrGlyAspLysValTh 334
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
709 .....GGATCAGGTGGTGGCCAG 727
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 rLeuLysThrGlyAlaValIleAspLeuSerGlyLysGluGlyGluIut 351
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
728 TCACCTTAGTAGTGAATAAATTAACATAGCCCATATGCTTT..... 771
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
351 hrTyrlLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeu 367
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
772 .....TTACCAACAGAGGCTCATTTGGCAGACGTGCTC 806
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
368 AlaLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGly.. 383
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
807 ACCAATGTTATCTATGATGCCCAAGCAAAAG..... 840
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
384 .....LysGluLysGlyGlyArgAlaIleVal 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
841 ..TGG.....TTAATTAATGGGGTATTGCAACG...GGCAAC 873
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
392 alTrpGlyAspIleAlaLeuIleAsnGlyAsnIleAsnAlaGlnGlySer 408
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
874 CCTATATAGAAAAGCAATGCTTC..... 900
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
409 GlyAspIleAlaLysThrGlyGlyPheValGluThrSerGlyHisTyrl 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
901 .....CAGCTGGTTCG...AAGATTTGTTTCATGATG 931
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 uSerIleGluSerAsnAlaIleValLysThrLysGluTrpLeuLeuAsp 442
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
932 AATCTTTGCTGAGATACCCATTCAGTATTCTACGAACGACGTCAAAAT 981
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
442 ro.....AspAspValThrIleGluAlaGluAspProLeuArg 454
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
982 GGGAAATCTCTTTTAAGCAGAT...AATAATGGCAAGGAAATCAAA 1028
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
455 AsnAsnThrGlyIleAsnAspGluPheProThrGlyThrGlyLualaSe 471
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1029 TGCCAAACATGAAACAATTCCTGCTCAATAGATTAAACAGCAACG 1078
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 rAspProLysLysAsnSerGluLeuLysThrThrLeuThrAsnThrIle 488
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1079 TTCAAATGCTTT.....AATGTTCTTTATCCGAG 1107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
488 leSerAsnThrLysLysAsnAlaTrpThrMetAsnIleThrAlaSerArg 504
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1108 ACAGCAAGCAACCTGTTATCATGCTGACGGTGTCAACAGATTATCG 1157
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 LysLeuThr.....ValAsnSer..... 510
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1158 ACCCAGACTGAATTAATGAGAAATATTTCTTATTTAGCAAGGAAAG 1207
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 ....SerIleAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysG 526
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1208 GCGAATGTACTTACCAACCAATCATCAATGCTGCGAGGAGATTATAT 1257
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 ly.....GlnArgIleGlyGlyValGln 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1258 TTCCAAGGAGATTTTACGGTCTCGCCTGAAATAATACGAAACTTGGCAAG 1307
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
534 IleAspGlyAspIleThrSerLysGlyAsnLeuThrIleTyrSerGlu 550

```

```

1308 GCGGGCGCTTCATATCATGTAAGACAGTACCCTTACTTGGAAAGTAAACG 1357
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
550 yGlyTrpValAspValIleHisLysAsnIleThrLeu..... 561
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1358 GCGTGGCAAAACGACCGCTGTCCAAATTCGCAAAAGCAACGCTGCACGTT 1407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
561 ..... 561
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1408 CAAGCCAAAGGGGAAACCAAGGCTCGATCAGCCGTGGCGACGATACGT 1457
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
562 .....AspGlnIlePheLeuAsnIleThrAlaIleSerVa 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1458 CATTTTGCATCAGCAGCAGACGATTAAGCAAAAAACACCTTTAGTG 1507
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
573 lAlaPhe...GluGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaL 589
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1508 AATATGCTTGGTGTACGCGGAGGGTACGGTCAACTGAATGCCATAT 1557
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
589 ysIle.....ValAlaGlnGlyThrValThrIleThrGlyGluGly 602
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1558 CAGTTCACCCCGCAAAACTATTTCCGCTTCGCGCGGACGTTTGA 1607
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
603 Lys.....AspPheArgAlaAsnAsnValSe 611
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1608 TTTAAACGGGCAATCGCTTTCGTCACCGTATTCAAAATACCATGAG 1657
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
611 rLeuAsnGly.....ThrGlyLysG 618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1658 GGGCGCATGTTGTC.....AACCAATACAGACCAAAAGATTC 1695
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
618 LysLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerGly 634
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1696 ACCGTTACCATTTACAGGCAATTAAGATTTGCTACAAACCGGCAATTAACA 1745
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
635 ThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThr..... 647
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1746 CAGCTTGATAGCAAAAAGAAATTCCTACCAACGTTGCTTGGCGAGA 1795
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
648 .....ThrArgLysAsnThrSerTyr.....TrpGlnThrSerH 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1796 AAGATACG.....ACCAAAACGAACGGGCGG 1821
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
659 lAspSerHisTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsn 675
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1822 CTCACCTGTTTACGAGCCCGCAGAAAGCCGACCTGCTGCTT.. 1869
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
676 PheThrPheIleLysTyrlleSerSerAsnSerLysGlyLeuThrThrGlu 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1870 .....TCGGGGGAAACAATTTAAACGGCACATCACGCAACAA 1909
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
692 nTyrlArgSerSerAlaGlyValAsnPheAsnGly.....ValA 705
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
705 snGlyAsnMetSerPhe..... 710
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1910 ACGGCAAACTGTTTTCAGCGGCGAGACCAACACCGGACGCTTAATCAT 1959
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1960 TTAACGACCATGTTGTCGCAAAAAGAGGCAATTCCTCGGGGGAATTCGT 2009
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
711 .....AsnLeuLysGluGlyAlaLysValAsnPheLysLe 722
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2010 GTGGACACAGCTGATCAGCCGACCA.....TTTA 2041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 ulysProAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheL 739
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2042 AAGCGGAAACTTCCAATTTAAAGCGGAGACGGCGTGT..... 2082
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
739 euAla...AsnIleThrAlaThrGlyGlyGlySerValPhePheAspIle 754
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2083 .....TCCGCGCATGTTGCCAAGTGAAA...GGCGATTGGCA 2117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
755 TyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSerIleLys 771
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2118 TTTGAGCAATCACGCCCAACAGCTTTTGGTGTCCGACCGCATCAAGCC 2167

```

771 nileserasnlglyAlaasnphetrleu.....AsnserH 783
2168 ACACATCTGTACACGTTCCGAGTGGACGGCTGCACAAATTGTGTGAA 2217
783 is.....ValArgglyAspaspAlaPhelysileasn..LysAsp 795
2218 AAAACCATTACCGAGCATAAAGTATGCTTCATCTGACTAGACCGCAT 2267
796 LeuthrileasnAlaThrAsnSerAsnPheserleuArglnThrLysAs 812
2268 CACGGGCAATGTCATCTTGCCGATCAACGCTCATTAATCTCACAGGC 2317
812 pAspPhetrlyaspGlyTyrAlaArgAsnAlaIleasnSerThrTyrAsnI 829
2318 TTGGCACACGTACAGGCAATCTTAGTGCAAATGGCGATACACGTTATACA 2367
829 IeserIleleuglyGlyAsnValThrleuIleugly..... 840
2368 GTCAAGCCACAAACGCCACCAAAAGCGCAACCTTAGCCTGGGGCAATGC 2417
841GlnAsnSerSerSerSerIlethrGlyAsn.. 850
2418 CCAGCAACATTTATCAAGC.....ACATTAACGGCAACACAT 2458
851IleThrIleGlyLysAlaAlaasnValThrleuGlyAlaAsnAsn 866
2459 CGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACACCGCGGTACAAAC 2508
866 LApraasnGlnIn.....AsnIleArgaspArgValIleIelysLeu 879
2509 GGCAGCTGACGCTTTCGGGCAACGCTTAAGCGCAACGTAAACCATTCGC 2558
880 GlySerleuLeuValasnGlySerleuSerleuThrGlyLysAsnAlaAs 896
2559 ACTCAACGCTAATGTCCTCCAGCCGATAGGACAGTATTCAT..... 2601
896 PiletysglyAsnleuthrIleSerGlnSerIleThrPhelysGlyLysT 913
2602 ..TTTGAACACAGCCGCTTACCGCAAAATACGCGCGGCAAGATAGC 2649
913 hrArgaspThrleuAsnIleThrGlyAsnPhetrAsnAsnGlyThrAla 929
2650 GCATTAACACTTAAAGACAGCAATGAGCCGTCGCCGACGGCGCAAGATT 2699
930 GlnIleasnIlethrGlnGlyValIleValIleuglyAsnValThrAsnAs 946
2700 AGGCATTTAAACCTTGACACGCCACCATTTACACTCAATTCGCCCTATC 2749
946 pGlyAspLeuAsnIleThr..... 953
2750 GCCACGATGGCGAGGGGCCAAACGGCGAGTCGACAGATGGCGCGC 2799
953 953
2800 CGCGCTTCGGCGCGCTTCGGCGCTTCCTATTATCCGT...TACACCGC 2845
954 ..HisAlaIlyAsArgasnGlnArgSerIleIleGlyAspIleIleAs 969
2846 CAACCTTCGGTAGAGATCCGTTTACACACGCTGACGCTAAACGGCAATTC 2895
969 nlysllysglySerleuAsnIleThrAspSerAsnAsnAspAlaGlnIleG 986
2896 AACGCTGAGGAAACATTCGCTTTATGTCGAACTTTCGGCTACGCGAG 2945
986 IlnleGlyGlyAsnIleSerGlnLysGlnGlyAsnleuthrIle..SerSe 1002
2946 CGACAAATGAAAGCTGGCGAAAGTTTCGAGAGCACTTACACCTTGGCGG 2995
1002 rAspIlyslleAsnIleThr..... 1008
2996 TCACAAATACCGGCAAGCAACCTGCAAGCGCTGCACAAATTCACGGTAGTG 3045
..... 3045

404 ThrPheValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu..... 417
1674 CCACATCAAGACAAAGATCCACCGCTTACCAT..... 1707
418 ...AsnIleAspAlaIleGlyIleAsnValThrValAsnGlySerPheA 433
1708ACAGCAATTAAGATATGCTACACCGCAAT 1740
433 IagIleIleIleGlyIleGlyIleGlyIleLeuThrIleAlaGlnAsnGly... 448
1741 AACACAGCTTGATAGTACCAAAAGAAATGCTTACACGCTTGCTGG 1790
449 SerTyrValLeuAlaGlyIleGlyIleSerMetAlaLeuThrGly..... 462
1791 CGAAGAAATACGACCAAAAGACGCGCTCAACCTTGTTTACACG 1840
463AspIleValValAspAspGlyAlaValLeuSerLeuGlyA 477
1841 CCGCCGCA.....GAGACCGCACCGCTGCTGCTT 1869
477 sPaIAlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeu 493
1870 TCCGCGCGAACA...AATTAAAGCGCACATCCGCAACAAACGCGCA 1916
494 AsnGlyGlyValLeuAspLeuSerAspPheSerThrTrpGlnSerGly 510
1917 ACTGTTTTCAGCGCAGACACACACCGCACCGCTTAC..... 1953
510 rSerTyrAsnAspGlyLeuGlyValSerIleSerSerGlyThrValIleG 527
1954AATCATTTAAACGCCAT 1971
527 lYSerGlnAspValValAspLeuAlaGlyIAspAsnLeuHisIleGly 543
1972 TGCTCGCAAAAGAGGCGCAT.....CTCG 1997
544 GlYAspGlyLysAspGlyValIleValValAspAlaSerAspGlyI 560
1998 CCGGGAATCGTGTGGACACAGCATGATCAACCGCATTTAAAGCG 2047
560 nValSerLeuAlaAsnAsnSerTyrLeuGlyThrThr..... 573
2048 AAACCTTCCAATTAAGCGCGACAGCGGCTTCCCGCAATGTGGC 2097
574GlnIleAlaSerGlyThrLeuMetValSerAspAsn..... 585
2098 AAGTGAAGCGCATTTGCGATTTGACCAATCACGCCCAAGCGTTTGG 2147
586 SerGlnLeuGlyAspThrHisTyrAsnArg.....GlnValIlePheThr 600
2148 TGTCCGACCGCATCAAGC.....CACACAATCTGTACAC 2182
600 rAspLysGlnGlnGlnSerValMetGluIleThrSerAspValAspThrA 617
2183 GTTGGACGAGCGGCTGTGACAAATGTGTGCAAAAAACCATTAACGAC 2232
617 rGSerAspAlaIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 631
2233 GATAAGATGATGCTTCAATGCTAAGACGCGCATCAGCGGCAATGTGCA 2282
632 AspGlyIleValAlaValAspAlaGlyValAspThrGlnTrpGlyAlaLe 648
2283 TCTTCCCATCAC.....GCTCATTTAAATC 2308
648 uMetLAspSerSerGlyGlnHisGlnAspGluGlySerThrLeuThrL 665
2309 TCACAGGCTTGCCCACTCAACGCGCAATGTAGCAAAATGGGAGATACA 2358
665 ystHisGlyAlaGlyThrLeu.....GlnLeuThrAlaSerGlyThrThr 679
2359 CGTTATATACGTGACGACCAACGCCAACCAAAAGCGCAACTTACCTGCT 2408
680 GlnSerAlaVal..... 683
2409 GGGCAATGCCCAAGCACATTTAATCAAGCCCATTAACGCGCAACACAT 2458
684ArgValGlnGlyGlyThrLeuLysGlyAspValA 695
2459 CCGCTTCGGGCAATGCTTCAATTAATCTAAGCGCACCGCC..... 2499
695 la.....AspIleLeuProTyrAlaSerSerLeu 704
2500 ...GTACAAACGCGAGTGTGACGCTTCCGCAACGCTAAGCGCAACGT 2546
705 TrpValGlyAspGlyAlaThrPheValThrGlyAlaAspGlnAspIleG 721
2547 AAGCATTCGCGACATC...AACGTAATGTCTCCCTACCGCATTAAGCGAG 2593
721 nSerIleAspAlaIleSerSerGlyThrIleAspIleSerAspGlyThrV 738
2594 TATTCATTTGAAGCAGCGGCTTACCGGCAACAATACGCGCGGCAAG 2643
738 alLeu.....ArgLeuThrGlyGln..... 744
2644 GATACGCGATTACACTTAAAGACAGC.....GAATGACGCT 2681
745 AspThrSerValAlaLeuAsnAlaSerLeuPheAsnGlyAspGlyThrLe 761
2682 GCCGTACAGCAGGANA.....TTAGCAATTTAAAC..... 2712
761 uValAsnAlaThrAspGlyValThrLeuThrGlyIleuAsnThrAsnL 778
2713CTTGACACGCGCACCATTAACCTCAAT 2739
778 euGluThrAspSerLeuThrTyrIleuSerAsnValThrValAsnGlyAsn 794
2740 TCCGCTTATGCGCACGATCGCGAGGCGGCAACCGCATGCTGCGACAGA 2789
795 LeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValAlaGlyAs 811
2790 TGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCCTTATTCGTTA 2839
811 p..... 811
2840 CACGCGCAACTTCGTTGAATCCGTTTCAACAGCTGACGGAACGCGC 2889
812ThrLeuThrValAsnGly 817
2890 AAATTAACGCTCAGGACATTCGCTTATGTGCGCACTTTCGCGC... 2937
818 AspTyrThrGlyGlyGlyThrLeuLeuLeuAspSerGlyLeuAsnGlyAs 834
2938 ...TACCGCAGCAGCAAAATTTGAAGCTGCGGGAAGTTCCGAAGGCACTT 2983
834 pAspSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsnT 851
2984 ACACCTTGGGCGTCAACAAT...ACGCGC...AAGCAACCTGCAAGCTC 3027
851 hrThrValValAlaAsnSerIleThrGlyIleGlyGlnProThrSerThr 867
3028 GAACAATTTGACGCTAGTGGAGAGAAAAGACACAACACCGCTGTCCGAAA 3077
868 GlyIleLysValValAspPheAlaAlaAspProThrGlnPheGlnAsnAs 884
3078 CCTTAATTTCAACCTGCAAAAGAA...CACGTGATGGCGGCGGCTGGC 3124
884 nAlaGlnPheSerLeuAlaGlySerGlyTyrValAsnMetGlyAlaTyrA 901
3125 GTTACCAACTCATCCGCAAGACGCGAGTTCGCGCTCATATATCCGCGT 3174
901 sPtyrThrLeuValGluAspAsnAsnAspTyrIleu..... 913
3175 AAGAACAAGAGCTTTCGACAAACTCGCAGAGCAAGCAAGCAAAAAACA 3224
913 913

```

3225 GGGGAAAAAGACACGCGCAAGCCTTGACGGCTGATTGCGCGCGGCG 3274
913 .....
3275 GCGATCGCGTGCAGAAAGACAGAAAGCGTTGCCAGACCGCGCGCGCAGCA 3324
914 .....
3325 GCGGCGGAAATGTCGGCATTATGACGCGGAGAGAAAGAAAGCGGT 3374
922 .....
3375 GCAGCGGATTAAGACACCGCCTTGCGGAAACGCGGAAAGCGGAAACCC 3424
923 ..ProProAspProAsp.....
3425 GCGCGGCTACACGCGCTTCCCGCGCGCGCGCGCGCGGATTTG 3474
928 ..ProthrProAsp.....
3475 CCGCACTGCAACCCCAACGCGACCCCAACGCGGCGGACCTGATCAG 3524
932 ProAspProthrProAspProAspProthrProAspProthrProthr 948
3525 CCGTTATGCCAATAGCGGTTGAGTGAATTTCCGCGACGCTCAACAGCG 3574
948 oAlatYrGlInProValLeuAsnAlaLysValGlyLTYrLeuAsnAsnL 965
3575 TTTTCGCGGTACAGGACGATTAAGACGCGCTATTTCGCGAAGACGCGCG 3624
965 euAlaGAlaAla.....
3625 AACGCGGTTTGACACGCGC.....
978 AsphIsAlaGlyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyG 994
3654 CACCAAAACAGCGCTTGCGCAAGATTTCCGCGACCTACCGCAACAAACCG 3703
994 yAspTYrHisTYrThrAlaAlaGlyGlnLeuAla.....
3704 ACCTGCGCAATCGGTATGCGAAGAAACCTGCGGACGCGCGCGC..... 3747
1009 sPTThrSerThrValGlnLeuSerGlyAspLeuPhSerGlyArgTYrPly 1025
3748 .....
1026 ThrAspGlyGlnTYrMetLeuGlyLeValGlyGlyTYrSerAspAsnG 1042
3774 GACCGAAACACCTTGACGACGCGCATCGCACTGCGACGCGCTTGCC 3823
1042 nGlyAspSerArgSerAsnMetThrGlyThrArgAlaAspAsnGlnAsn 1059
3824 ACGCGCGCGCTTTGCGG.....
1059 IsGlyTYrAlaValAlaGlyLeuThrSerSerTYrPheGlnHisGlyAsnG 1075
3856 AGGTTCTCATCGGCATCGACGCGCGCGCGGCTTTAGC.....
1076 LysGlnGlyAlaTYrPheAspSerTYrPheGlnTYrAlaTYrPheSerAs 1092
3900 CAGCCTTTCAGACGCGCATCGAGCAAAATCCGCGCGCTGCTGATTT 3949
1092 nAspValSerGlnGlnGlyAspGly.....
3950 AC.....
1104 yHisSerSerGlyIleIleAlaSerLeuGlnAlaGlyLTYrGlnTYrPhe 1120
3982 .....
1121 ProGlyArgGlyValIleGlnProGlnAlaGlnValIleTYrGlnG 1137

```

```

4026 GTCGCAAAAGCGATTAACCGCTACGAAAC.....
1137 yValGlnGlnAspPheThrAlaAlaAsnArgAlaArgValSerGln 1154
4057 .....
1154 erGlnGlyAspAspIleGlnThr.....
4099 GCGGCGATTAAAGCAGATTATTCATTCAAACCGCGCAACATTTCCAT 4148
1163 LeuGlyLeuHisSerGlyTYr.....
4149 CAGCCTTATTGAGCTGTCC.....
1175 lIleProthrLeuAspLeuAsnTYrTYrHisAspProHisSerThrGln 1192
4196 TCCGAACACGCGCTCAATACCGCGCTATTGGCTCAGGATTTCCGCAAC 4245
1192 lGlnGlnAspGlySerThr.....
4246 CGC.....
1207 ArgGlyGlnIleIleLysValGlyValThrGlyAsnIleSerGln...ArgVa 1222
4293 GTCCTTCACGCTGCCCGCGC.....
1222 lSerLeuArgGlySerValAlaTYrPheGlnLysGlySerAspPheAlaG 1239
4337 AACACGCGCGGCGCAATTAAGCTACCGCTGG 4371
1239 lInThrAlaGlyPheLeuSerMetThrValLysTYr 1250
seq_name: p1r2:H98323

```

```

seq_documentation_block:
Hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: H98323
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H98323
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1341 <KOR>
A:Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:gl5160106; GSPDB:GN00170
A:Gene: AGR_L_3085
A:Map position: linear chromosome

```

```

alignment_scores:
Quality: 278.00 Length: 1423
Ratio: 0.470 Gaps: 68
Percent Similarity: 41.532 Percent Identity: 20.239

```

```

alignment_block:
US-09-303-518D-649 x H98323 ..

```

```

Align seq 1/1 to: H98323 from: 1 to: 1341
688 GGCATATCCTTTGCACAAATGATCAGTGGTGGCAGCTACACTTAGG 737
1111:.....
128 GlyThrLeuValValGlnGlyGlyLeuGlnThrAsnSerValIleLeuG 144
738 TAGTGAATAAATTAACATAGCCCATATGCTTTTACCAACAGGAGGCT 787
144 yAla.....
788 CATTTGGCAGACAGTGGCTACCAATGTTTATCTATGATGCCAAAGCAA 837

```

149 er.....SerGlyThrAlaSerPheIle.....GlyAla 158
149 er.....SerGlyThrAlaSerPheIle.....GlyAla 158
838 AAGTGTAAATTAATGGGATTCACAAACGGC..... 870
159 AsnTrpAlaAsnAspGlyAspValSerValGlyGlyAsnGlyThrGlySer 175
871AACCCCATATAG 883
175 rLeuLeuLeuHisGlyThrThrAlaSerSerGluAsnIleTyrValG 192
884 GAAAGAGC...AATGCTCCAGCTGGTTCGTAACAT..... 918
192 LysIleSerGlyThrGlySerGlySerLeuLysLeuAspAsnSerThr 208
919 ...TGTTCTATGATGAAATCTTGTGAGATACCAT..... 954
209 LeuThrValAlaAspThrValPheAlaGlyTyrAsnSerTrpSerSer 225
955TCAGTATCT 964
225 rAlaAlaGlyGlySerGlyAsnValGluValLeuGlyAlaSerSerLeu 242
965 ACGACACAGCTCAAAATGGGAATCTCTTTACGACGATTAATGCG 1014
242 ySer.....AlaAsnGlyValLeuAlaAsnAspProAspThrValGly 256
1015 ACAGGAAATCAATGCAACATGCAACATCTCTGCGTAATGATGAT 1064
257 ThrAlaLeuValSerGlyGlySerSerTrpAsnMetSerSerSer 273
1065 A.....AAAACAGAACGTTCAATGTTTAAATG 1093
273 uValValGlyGlyAlaGlyValGlyThrLeuThrIleThrGlyGlyM 290
1094 TTCTTTATCCGAGACAGAGAACGCTTTATCATGCTGACAGTGT 1143
290 eValIleSerGlnThr..... 295
1144 GTCAACAGTTATCCACCACTGATTAATGAGAAATATTTCCCTTAT 1193
296 ...ThrThrValIleAlaAspLeuAsnSerAlaAspGlySerSerValTh 311
1194 TGACGAGAGAAAGGCGAATGATCTTACGACCAATC.....A 1234
311 rValSerAlaGlyLeuGlySerSerLeuLysSerSerLeuValGlyA 328
1235 ATCAAGCTGCTGAGAGATTAATTCACAGAGATTTTACGCTCGGCT 1284
328 snGlyValAlaAlaLysLeuValValGluAlaAlaIleThr... 343
1285 GAAATTAACGAACCTTGGCAAGC.....GCGGGCTTCATATCAG 1325
344SerGlyIleAlaIleIleGlyAlaGlnHisSerAl 354
1326 TGAAGACAGTACCTGTTACT.....TGG..... 1347
354 aserGluAlaThrValIleThrGlyAspGlySerLysTrpThrThrGlyAspL 371
1348 ...AAAGTAAACGCGCTGGCAACGACGCTGTCCAAATCGGCAAGGC 1395
371 euGlnValGlyGlyAspThrSerAspProGlyGlyLeuAlaGlyAsnIly 387
1396 ACGTGCAGCTCAAGCCAAAGGGGA.....AACCA 1427
388 ThrLeuAsnValThrAlaGlyLysSerValAspSerThrValAlaHisIle 404
1428 AGGCTGCATCAGCGTGGCGAGCATCATTTGGATCAGCAGGAG 1477
404 uGlyValValAlaGlyAlaThrGlySerAlaIleValAsp..... 417
1478 ACGATAAGCGCAAAACACAGCTTTAGT.....CAA 1509
1478 ACGATAAGCGCAAAACACAGCTTTAGT.....CAA 1509
418GlyLysGlySerValTrpThrValAlaAspArgAsnSerLeuGlu 431
1510 ATCGGCTGTGTCAGCGAGGGGTACGTACGTCACTGAATGCCGATATCA 1559
432 ValGly...ValSerGlyAlaGlySerLeuAlaValThr..... 443
1560 GTTCAACCCCGACAAACTATTTGGCTTTCGCGGAGCGTTTGAT 1609
444GlyGlyGlyLeu.... 447
1610 TAAACGGCATTCGCTTGTTCACACCGATTCAAATACGATGAAGG 1659
448ValAspAlaAla 451
1660 GCGATGATTCACACCAATCAAGACAAGAAATCCACGTTACCATTC 1709
452 AsnIleIleIleGlyThrAsnThrGly..... 460
1710 AGCAATTAAGATATTGCTACACCGGCATTAACAGCTTGGATGCA 1759
461 .GlyAsnGlySerValArgValSerGlyAlaAspSerThrValLysSerA 477
1760 AAAAGAAATTCCTCAACAACGCTTGTGGCGAGAGAATACGACCAA 1809
477 rGSerAspLeuAsnVal...GlyLeuTyrGly..... 486
1810 ACGAACGGCGGCTCAACCTGTTTACCAAGCCCGCGAGAACCGC... 1857
487 ...AsnGlySerMetThrValGluAlaGlyAlaValLysSerArgAs 502
1857 1857
502 pGlyTyrValAlaThrTyrGlyGlySerThrSerAlaValThrValThre 519
1858ACCTGCTGCTTTCGCGCGA 1878
519 LysAspGlySerSerTrpAlaMetThrGlyThrPhePheValGlyTyrAla 535
1879 ACAATTTAAACGCGACATCAGCAACAAACGCAACCTGTTTTCAG 1928
536 SerGlyAlaThrGlyAsnValThrValSerAsnGlyAlaIle... 550
1929 CGGACAGCAACACCGCACCGCTTCAATTCATTAAGACCATGTGCGC 1978
551A 551
1979 AAAAGAGGCGATTCCTCGCGGGAATCGTGGC..... 2013
551 rGAlaThrGlyValThrLeuGlyAspLeuAlaGlyAlaSerGlyThrMet 567
2014GACACGAGCTG 2024
568 ThrIleThrGlyAlaGlySerLysValThrAlaTyrValAspAsnGlyTh 584
2025 GATCAACCGCACATTTAAACGGAACCTTCAATTAAGCGGAGCAGG 2074
584 rValAsnSerGlySerValAspValGlyPheGlnGlySerGlySerLeuS 601
2075 CGGAGTTCCTCCGCAAT.....GTTGCCAAA 2100
601 eValValAsnGlyGlySerLeuAspAlaTyrAsnLeuTyrValGlyAsn 617
2101 GTGAAGCGCATTCGCTTACCAATCAGCAACGCAACGATTTTGGTG 2150
618 AlaLeuGly.....SerSerGlyAlaValLeuValSerGlyVal 630
2151 CGCACCGCATCAAC..... 2166
630 lGlySerHisValSerValAspGlyLeuMetValValGlyAsnAlaGlyA 647
2167CAC 2169
647 snGlySerValGluIleThrGlyGlyAlaSerLeuAlaIleProThrIle 663

2170 ACAATCTGACAGCTTCGACGAGCGGTCTG..... 2202
|||:|||||:|
664 LeuLeuLeuThrGluAlaGlySerThrGlyValLeuSerIleGlyAlaGly 680
2203ACAAATGTGTGCGAA...AAAACCATTA 2227
|||:|||||:|
680 ySerGlyGlnThrAlaArgSerAlaGlyAlaValAlaArgAlaLeu 697
2228 CC.....GACGATAAAGTATGCTTCATTGACTAGAAC... 2262
|||:|||||:|
697 IapheGlyAlaGlyAsnGlySerIleValPheAsnHisSerGluThrGly 713
2263GACATCAGCGGCAATGCGATCTTCCGATCAGC 2297
|||:|||||:|
714 TyrThrLeuSerAlaAspIleSerGlyAlaGlyArgValAlaAlaGlu 730
2298 TCATTTAAATCTCAGAGGCTTGCACACTCAAGCAATCTTAGTCAAA 2347
|||:|||||:|
730 a.....GlyValThrThrLeuSerGlyAsnAsnSerTyr 742
2348 ATGGGATACACGTTATACA.....GTACGCCACAAAGCCACCA 2388
|||:|||||:|
742 erGlyGlyThrThrIleSerAlaGlyMetLeuLysGlyThrAlaLysSer 758
2389 AACGGCAACCTTAGCTCTGCGGCAATGCCCAAGCAACATTATATCAAGC 2438
|||:|||||:|
759 PheGlySerGlyGlyIleValAsnAsnAlaGluLeuValAlaAsp.... 773
2439 CACATTTAAAGCGGCAACACA.....TCGCTTCGGGCAATGCTTCAT 2479
|||:|||||:|
774GlyGlyGlyThrLeuSerAsnAlaIleSerGlyThrGlySerP 788
2480 TTAATCTAAGGAGCCAGCCGCTACAAAAGCAGCTGACGCTTTCGCGC 2529
|||:|||||:|
788 heGluLysThrGly.....AspGlyAsnLeuLeuLeuThrGly 800
2530 AACGGTAAGGCAAC.....GTAGCATTCGCGACTCAACGG 2567
|||:|||||:|
801 AsnSerThrTyrSerGlyAlaThrAlaValSerAlaGlyLysLeuSerVal 817
2568 TAAATGTCCTCCAGCCGATTAAGCAGTATTCATTTTGAAGCAGCCGCT 2617
|||:|||||:|
817 LAsnGlySerIleuAlaSerAlaVal.....SerValG 828
2618 TTACGGGACAATTCAGCGCGGCAAGATAGCGCATTACACTTAAA... 2664
|||:|||||:|
828 LysSerGlyAlaThrValGlyGlyThrGlyThrIleGlyGlyLeuThrVal 844
2665 GACAGGAAATGAGCGCTGCCGTCAGGACGAGATTAAGCAATTTA...AA 2711
|||:|||||:|
845 AsnSerGlyGlyThrLeuAlaProGlyAsnSerIleGlyThrLeuThrSe 861
2712 CCTTGACAACGCCACATTACACTCAATTCGCGCTAT.....CGCCACG 2755
|||:|||||:|
861 rThrGlyAsnAlaThrPheAlaSerGlySerThrTyrAlaValGluLeu 878
2756 ATGCGGCAAGG.....GCCCAACGGGAGTCGCGACAGAT 2790
|||:|||||:|
878 sPAlaSpIleYerSerAspArgLeuAlaValThrGlyThrIleThrIle 894
2791 GGGCGCGCGCGCGGTTCGCGCGCTTCCTTATTCGCTTAC 2840
|||:|||||:|
895 Ala.....AsnAspValSerLeuIleValAla 903
2841 ACCG.....CSAACTTCGGTAGAATCCGTTTCAACAGCGC 2875
|||:|||||:|
903 rProLeuGlyAlaHisSerAlaTyrSerLeuGlyThrArgTyrThrIleL 920
2876 TAGCGTAAACGGCAATTCAGACGTCGAGGAACATTCGCTTATGTCG 2925
|||:|||||:|
920 euThrAlaThrGlyGlyValThr.....GlyThrPheSerSerValAlaSp 934

2926 GAACCTTCGGCTACCGGACGACAAATTTGAAGCTGGCGGAAATTCGGA 2975
|||:|||||:|
935 GluSerPheAlaTyrLeuThrAlaLysValAlaGlnSerGlyAspAspAl 951
2976 AGGCACCTTAC.....ACCT 2989
|||:|||||:|
951 aThrThrTyrLeuSerPheLeuArgThrSerProAspSerGlyLeuLeu 968
2990 TGGCGGTCAACAATACCGGCAACGAACTGCAAGCTCGCAACATTTGACG 3039
|||:|||||:|
968 lAlaAlaThrSerThrAlaAsnAlaArgAlaAlaAlaAlaValAla 984
3040 GTAGTGAAGAGAAAGACAACAACCGCTGTCCGAAACCTTAATTTCAC 3089
|||:|||||:|
985 AlaLeu.....GlyGlnThrSerProLeuTyrGluAlaAlaLeuPheLe 999
3090 CTTGCAAAACGAAACAGCTGATGCCGCGCGGTGGGTTACCAACATCC 3139
|||:|||||:|
999 uGlnGlnGlyGlu.....ThrGlnGlyAlaPheSer...GlnLeuAla 1012
3140 GCMAAGACGCGAGTTCGCTGCATTAATCCGTCMAAGAACAGAGCTT 3189
|||:|||||:|
1013GlyGluIle.....HisProSerLeuAlaMetAlaLeu 1023
3190 TCCGACAAACTCGGCAAGCAGAGCAACCAAAAACGCGGAAAGACAA 3239
|||:|||||:|
1024 lLeAsnArgSerGly..... 1028
3240 CGGCAAAAGCTTAGCGCGCTGATTCGGCCGCGGCGCGATGCGTGANA 3289
|||:|||||:|
1029GlnSerArgAspValAlaIleLeuAsnArgLeuArgSerAlaPheGlu 1044
3290 AGACACAAGCGTTGCCGAAACCGCGCCGCGAGGACGCGGGGAAATGTC 3339
|||:|||||:|
1044 LyValAlaSp..... 1046
3340 GGCATTATTCAGCGCGAGAGAGAGAAAAAGGGTCCAGCGGATTAAGA 3389
1046 1046
3390 CACCGCTTGGCGAAMACAGCGGACGGAACCGCGCGCTTACCAACG 3439
1046 1046
3440 CTTTCCCGCGCGCGCGCGCGCGGATTTGCCGCACTGCACACC 3489
|||:|||||:|
1047AlaArgProIleLeuPro..... 1052
3490 CAACCGACGCCCAACCGGACGCGACCTGATCAGCGCTTATGCAATAG 3539
|||:|||||:|
1053ValAlaTyrAlaGluG 1058
3540 CGGTTAGTAGTAATTTCCGCCACGCTCAACAGCGTTTTCGCGTACAG 3589
|||:|||||:|
1058 yGlyGlnAsnProPheAla..... 1064
3590 ACGAATTAGACCGCGTATTTCCGAGAGACCGCGCAACGCCGTTTGACA 3639
|||:|||||:|
1065ValAspAspGlyAlaLeuSerPheThrPser 1074
3640 AGCGGCAATCCGGGACACCAACAACTACCGTTCCGAAATTTCCGCGCTA 3689
|||:|||||:|
1075 SerGlyPheArgSer..... 1079
3690 CCGCAACAACAAACGACCTGCGCAATCGTATGACAGAAACCTCGGCA 3739
|||:|||||:|
1080 ArgGlyArgIleAspSerAspGlyAsnGlySerSerValAspMetAsnG 1096
3740 GCGGCGCGCTGCGCATCTGTTTTCGCAACAACGAGCAACCAACACCTTC 3789
|||:|||||:|
1096 LysGly.....GlyIleLeuPheGly.....Leu 1103
3790 GACGACGCACTCGGCACTCGGACGCGCTTGCCACGCGCGCTTTCGG 3839

```

1104 AspGlyValLeuSerAspSerTrpArgAlaGlyValAlaAlaGlyIyrGI 1120
3840 GCATACGGCATCGACAGGTTTACATCCGACACAGCGGGGGCGGGT 3889
1120 YHLSAspGlyIleYsgInThrAlaLeuAlaAlaSerAla..... 1133
3890 TTAGCAGCGGCGACCTTTTCAGACGGCATCGAGGCAAAATCCGCGCGCG 3939
1133 ..... 1133.
3940 GTGCTGATTCATGCGCATTCAGCAGATACCGCGCGGTTTGGCGGAT 3989
1134 .....AspValAspSerTyrTyrLeuAlaAlaIatyrAlaGlyTh 1146
3990 CGGATCGAAGCG.....CACATCGGGCGCAACGGCGCTTTTCGTGC 4030
1146 rGInIleGlyProAlaSerLeuArgLeuGlyAlaIleHisAlaPhe...G 1162
4031 AAAAGCGGATTCACCGTACGAAACGTCAATATCGCCACCGCGCGCTT 4080
1162 LAspAlaGlyThrArg.....ArgSerIle 1170
4081 GCATTCACCGGTACCGCGCGGCGCATTCAGCAGATTCATTCAAACC 4130.
1171 SerPheSerThrLeuGlnIleuSerLeuSerAlaSerTyrGlyGlySerTh 1187
4131 GCGCGCAACATTTCC..... 4146
1187 rSerGlnValPheAlaGlyAlaIleTrpArgPheAspPheLeuThr 1204
4147 ..ATCAGCGCTTATTGAGCCGTCTCTATACCGATCGCGCTTCGGGCAAA 4194
1204 IsIleGlnProTyrAlaAsnIleAlaIatyr..... 1213
4195 GTCCGACACGCGTCAT.....ACCGCGTATTCGCTCA 4229
1214 ValAsnThrArgThrAspGlyPheGlnIleuYsgIlyAlaIleAlaIaVa 1230.
4230 GGATTCGGCAAAACCGCG 4248
1230 lSerSerGlySerThrArg 1236
seq_name: p1r2:S28634
seq_documentation_block:
addressin AIDA-I precursor - Escherichia coli plasmid pIB6
C:Species: Escherichia coli
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28634; S22680; S28881; S72657
R:Benz, I.
submitted to the EMBL Data Library, March 1992
A:Reference number: S28634
A:Accession: S28634
A:Molecule type: DNA
A:Residues: 1-1286 <BEN>
A:Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA46156.1; PID:g42255
Mol. Microbiol. 6, 1539-1546, 1992
A:Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escheri
A:Reference number: S22680; MOID:92326638
A:Accession: S22680
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 839-1286 <BE2>
A:Cross-references: EMBL:X65022
A:Experimental source: strain 2787
A:Accession: S28881
A:Molecule type: protein
A:Residues: 50-56 <BE3>
A:Experimental source: strain 2787
R:Suhr, M.; Benz, I.; Schmidt, M.A.
Mol. Microbiol. 22, 31-42, 1996

```

```

A:Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the out
A:Reference number: S72657; MOID:97055419
A:Accession: S72657
A:Molecule type: protein
A:Residues: 847-856 <SUN>
A:Experimental source: DAEC strain 2787
C:Genetics:
A:Genome: plasmid pIB6
C:Keywords: membrane protein
F:1-49/Domain: signal sequence #status predicted <Sig>
F:50-1286/Product: adhesin AIDA-I #status predicted <Mat>

alignment_scores:
Quality: 275.00 Length: 1458
Ratio: 0.434 Gaps: 65
Percent Similarity: 43.416 Percent Identity: 18.999

alignment_block:
US-09-303-518D-649 x S28634 ..

Align seg 1/1 to: S28634 from: 1 to: 1286
688 GCATACCTTTGCAAAATGATCAGT.....GGTGGCAC 725
|||||:||||| ||| |||||
45 GlyAsnAlaPheAlaValAsnIleSerGlyThrValSerSerGlyGlyTh 61
726 AGTCACTAGTAGTGAATAAATTAACATAGCCATATGT..... 768
|||||:||||| ||| |||||
61 lValSerSerGlyGlnThrGlnIleValTyrSerGlyArgGlyAsnSerA 78
769 .....TTTTACCAACAGAGGC... 786
78 snAlaThrValAsnSerGlyGlyThrGlnIleValAsnAsnGlyGlyLys 94
787 .....TCATTTGGCGGACAG 800
95 ThrThrAlaThrThrValAsnSerSerGlySerGlnAsnValGlyThrse 111
801 TGGCTGACCAATG.....TTTATCTATGATGCCCAAAACAAAGTGT 844
|||||:||||| ||| |||||
111 rGlyAlaThrIleSerThrIleValAsnSerGlyGlyIleGlnArgValS 128
845 TAAATTAAGGATTTGCAACGGCGACCCCTATATAGCAAAACAGAT 894
|||:||||| ||| |||||
128 rSerGlyGlyValAlaSerAlaThrAsnLeuSerGlyGlyAlaGlnAsn 144
895 GCGTTCCAGCTGCTGTAAGAATGGTCTATGATGAATCTTTCGTG 944
|||:||||| ||| |||||
145 lIleTyrAsnLeuGlyHisAlaSer.....AsnThrValIlePheSerGI 159
945 AGATACCATTCAGTATTC..... 963
159 YGlyAsnGlnThrIlePheSerGlyGlyIleThrAspSerThrAsnIleS 176
964 .....TAGAACACAGTCAAATGCGAAATCTTTTAAACGAC 1002
176 rSerGlyGlyGlnIleArgValSerSerGlyGlyValAlaIaSerAsnThr 192
1003 GATTAATATGCGACAGAAATCAATGCCCAACATGAACACAAATCTGT 1052
|||||:||||| ||| |||||
193 ThrIleAsnSerSerGlyAlaGlnAsnIleLeuSerGlyGlyAlaI 209
1053 GCCTAATAGATTAAACACAGAACCGTCAATGTTTAACTGTTCTTAT 1102
|||:||||| ||| |||||
209 rSerThrHisIleSerSerGlyGlyAsnGlnIyr.....IleSerAlaG 224
1103 CGAGACACAGACAGACCTGTTATCATGTCGACGGTGTGTCAACGT 1152
|||:||||| ||| |||||
224 lYalaSnAlaThrGlnThrIleValAsnSerGlyGly..... 236
1153 TATCGACCGAGCTGAATATGGA..... 1176

```

237 ... PheGlnArgValAsnSerGlyAlaValAlaIleThrGlyThrValLeuSe 252
1177 GAAATATTTCTTATTATGACGAGGAAAGCGGATTTGA 1216
252 rGlyGlyThrGlnAsnValSer SerGlyGlySerAlaI 265
1217 TACTTACGACACATCATCAAGTGTGAGGATTATATTTCCAGGA 1266
265 IeSerThrSerValTyrAsnSerGlyValGlnThrVal... PheAlaGly 280
1267 GATTTACGCTCTCGCTGCAAAATACGAAACTTGCA 1304
281 AlaThrValThrAspThrThrValAsnSerGlyGlyAsnGlnAsnIleSe 297
1305 AGCGCGGCGGCTTCATATCATGACAGACAGTACCTTACTTGAAAGTAA 1354
297 rSerGlyGlyIle... ValSerGlu... ThrThrValAsnValS 310
1355 ACGGCGGCGCAAAC GACGCGCTGTCCAAATTCGCG 1389
310 eRgLyThrGlnAsnIleTyrSerGlyGlySerAlaLeuSerAlaAsnIle 326
1390 AAAGGACGCTGCACCTTCAAGCCAAAGGGAA 1422
327 LysGlySerGlnIleValAsnSerGluGlyThrAlaIleAsnThrLeuVa 343
1423 AACCAAGCTCGATCAGCTGGCG 1447
343 lSerAspGlyGlyTyrGlnHisIleArgAsnGlyGlyIleAlaSerGly 359
1448 ACGGTACAGCATTTTGATCAGCAGCAGCATTAAGCAAAACAA 1497
360 ThrIleValAsnGlnSerGlyTyrValAsnIleSerSerGly 373
1498 GCCTTATGTAATCGCCTGTGTCAGCGGAGGCTACGCTGCAACTGAA 1547
374 GlyTyrAlaGlnSerThrIleIleAsnSerGlyGlyThrLeuArgValIle 390
1548 TGGCGATATTCAGTTCAACCCGCAAACTCTATTTCGGCTTCGCGCG 1597
390 uSerAspGlyTyrAlaIleArgGlyThrIleLeuAsn AsnSerG 404
1598 GACGTTTGATTTAAACGGGCAATCGCTTCGTTCCACCGTATTCAAAT 1647
404 lYArgGlnAsnValSerAsnGlyValSerTyrAsnAlaMetIleAsn 420
1648 ACCGATCAAGGCGGATGATTTGTCACACATCAAGCAAAAGATCCAC 1697
421 ThrGlyGlyAsnGlnTyrIleTyrSerAspGlyGlnAlaThrAlaAlaI 437
1698 CGTTACCATTCAGCAGCATTAAGATATTGTACACCGGCAT 1740
437 eValAsnThrSerGlyPheGlnArgIleAsnSerGlyGlyThrAlaProv 454
1741 .. AACACAGCTTGATAGCAAAAAAGAAATTGCC 1773
454 aGlnAsnSerValValAlaThrArgThrValSerSerAlaAlaLysPro 470
1774 TACAAGGT 1782
471 PheAspAlaGluValTyrSerGlyGlyLysGlnThrValTyrLeuThrPAr 487
1783 TGGTTTGGCGAAGAAATACG 1803
487 gGlyIleThrPtyrSerAsnPheLeuThrAlaValTyrSerMetPhePro 504
1804 ACCAAAGCAAGCGGCGGCTCAAC 1827
504 lYThrAlaSerGlyAlaAsnValAsnLeuSerGlyArgLeuAsnAlaPhe 520
1828 CTTGT 1832
521 AlaGlyAsnValValGlyThrIleLeuAsnGlnGlnGlyArgGlnTyrVa 537
1833 TTACAGCGCCGCGCA GAAGACCGCA 1858
537 lTyrSerGlyAlaThrAlaThrSerThrValGlyAsnAsnGlnGlyArg 554
1859 CCTGCTGCTTCCGCGGAAACAATTTAAACGGCAATCAGCAACACA 1908
554 lutyValLeuSerGlyGlyIleThr... AspGlyThrValLeuAsnSer 569
1909 AACGGCAACAGTCTTTCAGC GCGAGCACACACCGCAGGCTA 1952
570 GlyGlyLeuGlnAlaValSerSerGlyGlyLysAlaSerAlaThrValI 586
1953 CAATCATTTAAACGACCATTTGTCGCAAAAGAGGCAATTCCTCGCGGG 2002
586 eAsn GluGly GlyAlaG 592
2003 AAATCGTGTGGACACAGCAGTGCATCAACCCGACATTTAAAGCGGAAAC 2052
592 lnPheValTyrAspGlyGlyGlnValThrGlyThrAsnIleLysAsnGly 608
2053 TTCCAAATTTAAAGCGGACAGCGGCTGCTCCGCAATGTTGCC... AA 2099
609 GlyThrIleArgValAspSerGlyAlaSerAlaLeuAsnIleAlaLeuSe 625
2100 AGTGAAGCGCATTTGCATTTGACCAATCACGCC 2133
625 rSerGlyGlyAsnLeuPheThrSerThrGlyAlaThrLeuProGluLeu 642
2134 CAMCAATTTTGTGTGTCGACCG... CAMCAAGCACACACATC 2175
642 hThrMetAlaIleAsnSerValSerGlnAsnHisAlaSerAsnIleVal 658
2176 TGTACACGTTGCGACTGCGAGCGGCTGTGCAAAATTTGTGCAAAACCAT 2225
659 LeuGlnAsnGlyGlyLeuLeuArgValThrSer GlyGlyThrAl 673
2226 TACCGACGTAAGAATGATGTTCTTCACTAGCTAAGCCACATCAGCGCA 2275
673 aThrAsp ThrThrValAsnSerAlaGlyA 683
2276 ANTCGATGTCGCGATCAGCTCATTTAAATCTACAGCGGCTTCCCA 2325
683 rGluLeuArgIleAspAspGlyGlyThrIleAsn GlyThrThr 697
2326 CTCACGCGC 2334
698 lLeAsnAlaAspGlyIleValAlaGlyThrAsnIleGlnAsnAspGlyAs 714
2335 ATCTTAGTGCMAATGGCGATACAGTTTACAGTCAGCGC 2374
714 nPheIleLeuAsnLeuAlaGlnAsnTyrAspPheGlnThrGluLeuSerG 731
2375 ACACGCGC ACCCAAAACGCG 2394
731 lYserGlyValLeuValLysAspAsnThrGlyIleLeuThrThrAlaGly 747
2395 AACCTTAGCCTC GTGGCAATGCCCAAGCAACAT 2429
748 ThrLeuThrGlnAlaGlnGlyValAsnValLysAsnGlyGlyIleLeu 764
2430 TATATCAAGCACATTAACGCGCAACATCGGCTTCCGCAATGCTTCAT 2479
764 eAspSerAlaValAlaAsnAlaAspMetAlaValAsnGlnAsnAlaTyrI 781
2480 TTAATCTAAGCGACACGCGCGTACAAACGCGAGTGTGACGCTTCCGCG 2529
781 lAsnIleSerAspGlnAlaThrIleAsnGlySer ValAsnAsn 795
2530 AACGCTAAGCAACAGTAAAGCATTCGCCATCGCACTCAAGGTATGTCCT 2579
796 AsnGlySerIleValIleAsnAsnSerIleIleAsnGlyAsnIleThr... 811

[illegible]


```
1181 heAlaVal..... 1183
3730 AACCTCGCAGCGCGCTCGGCATCTGTTTCGACAAACCGGACCGA 3779
      ::::::::::::::
1184 .....LysGlyArgAlaAla..... 1188
3780 AAACACCTTGACAGCGCATCGCAACTGCGCAGCGCTTGCCACGCGG 3829
      ::::::::::::::
1189 ....AsnPheAspGlnLysLeu.....AsnGlyW 1197
3830 CCGTTTCGGGCAATACGCATCGACAGTTTACATCGGCATCAGCGCG 3879
      ::::::::::::::
1197 etrHLeuGlyGlyAspLysAlaThrAlaLeuAlaAspGlyValLeuSer 1213
3880 GCGCGGGGTTTATGACAGCGGCGAGCTTTCAGAC.....GG 3914
      ::::::::::::::
1214 ValGlyGlyPheAlaSerTyrSerSerSerAspIleYThrAspTyrG1 1230
3915 CATCGGAGGCAAAATCCGCCCGCGCTGTCATTCAGGCATTCAGGCAC 3964
      ::::::::::::::
1230 nSerLysGlyLysValAspSerHis.....SerPheGlyAlaTyrAlaG 1245
3965 GATACCGCGCGCGGTTTCGCGGATTCGCAATCGAACCGCACATCGGCGCA 4014
      ::::::::::::::
1245 InTyrLeuAlaAsnSerGly..... 1251
4015 ACGCGCTATTTC.....GTCCAAAAGCGGATTACCGGTACGAAA 4055
      ::::::::::::::
1252 .....TyrTyrMetAsnAlaValAlaValLysAsnAsnGlnPheSerGlnAs 1266
4056 CGTCAATTCGCCACCCCGCGCTTGCAATTCACCGCTACCGCGCGGCA 4105
      ::::::::::::::
1266 pValAsnIleThr.....SerIleAsnGlySerAlaSerGlyY 1279
4106 TTAAAG.....GCAGATTATTCATAAACCGCGCACACATTT 4143
      ::::::::::::::
1279 alSerAsnPheSerGlyMetGlyIleAlaLeuLysAlaGlyLysHisPhe 1295
4144 TCC.....ATCAGCGCTATTGTAGCGCTGTCCATTACCGA 4178
      ::::::::::::::
1296 AsnPheAsnGlnAlaTyrValSerProTyrValAlaMetSer.....Al 1310
4179 TGCCGCTTCGGGCAAAAGTCGACACAGCGGTCAATACCGCGTATTGGCTC 4228
      ::::::::::::::
1310 aPheSerSerGlyLysSerAsnIleSerLeuSerAsnGlyMetGlnAlaG 1327
4229 AGGATTTCGGCAAAACCGCGAGTGG.....GAATGGGCGGTAAAGCC 4272
      ::::::::::::::
1327 InSer...SerSerThrArgSerAlaMetGlyThrLeuGlyValAsnAla 1342
4273 GAAATCAAGGTTTCACGCTGCCCTCCACGCTCCGCGCCAAAGGCC 4322
      ::::::::::::::
1343 .....GlyTyrArgPheValMetAsnAsnGlyAlaGlnLeuLysPr 1356
4323 G 4323
      |
1356 o 1356
```